Линейная классификация и метрики качества классификации

Повторение

Среднеквадратичная ошибка

- MSE для линейной регрессии:

  $\frac{1}{l}||Xw - y||^2 \to \min\_w$

$$Q(w_1, \ldots, w_d) = \frac{1}{l}\sum_{i=1}^{l}(w_1 x_1 + \ldots + w_d x_d - y_i)^2$$

Можно посчитать градиент MSE:

- $\nabla \frac{1}{l}||Xw - y||^2 = \frac{2}{l}X^T(Xw - y)$

Приравниваем нулю и решаем систему линейных уравнений:

- $w = (X^T X)^{-1} X^T y$

Аналитическое решение

$$w = (X^T X)^{-1} X^T y$$

- Если матрица $(X^T X)$ вырожденная, то будут проблемы
- Даже если она почти вырожденная, все равно будут проблемы
- Если признаков много, то придется долго ждать

- Обращение матрицы – сложная операция ($O(N^3)$) от числа признаков

- Если заменить среднеквадратичный функционал на другой, то скорее всего не найдем аналитическое решение

Метод градиентного спуска

- Наша задача при обучении модели – найти такие веса w, на которых достигается минимум функции ошибки.

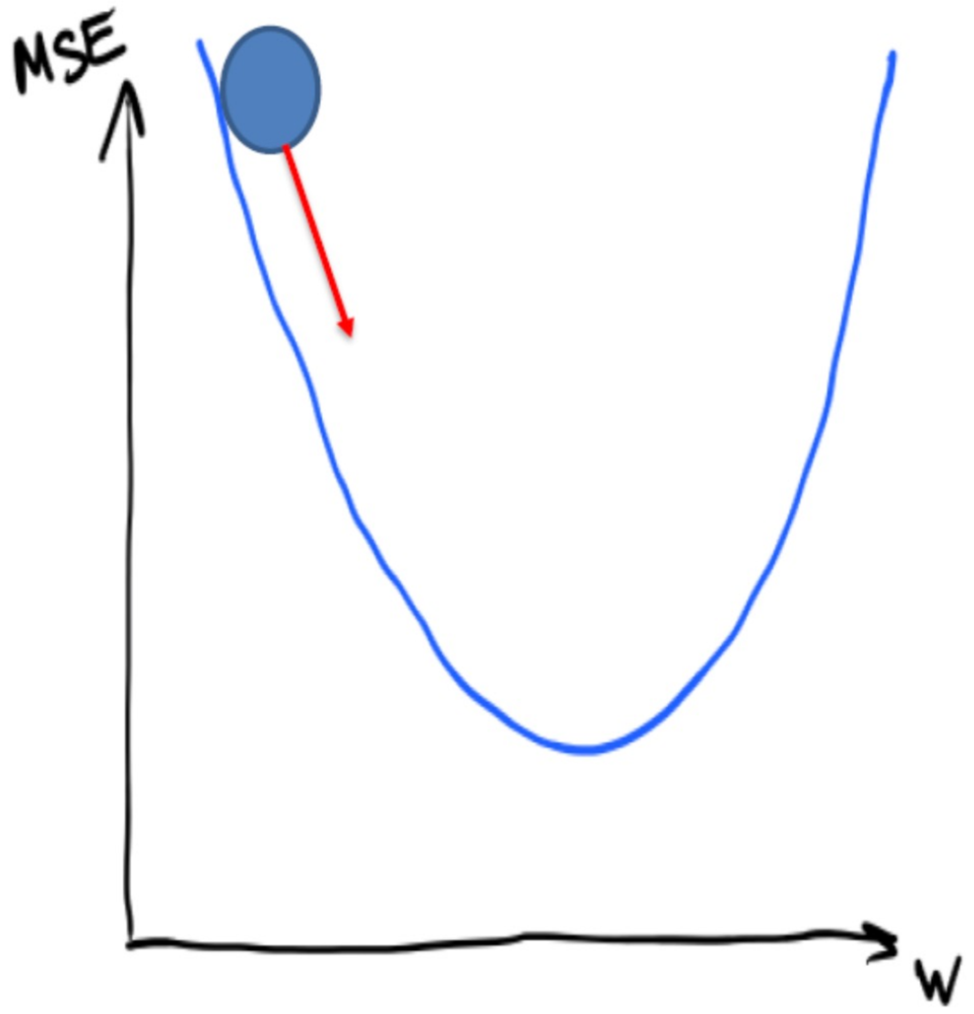
- Грубо говоря, график MSE – парабола

- Идея метода градиентного спуска.

- На каждом шаге движемся в сторону антиградиента функции потерь!

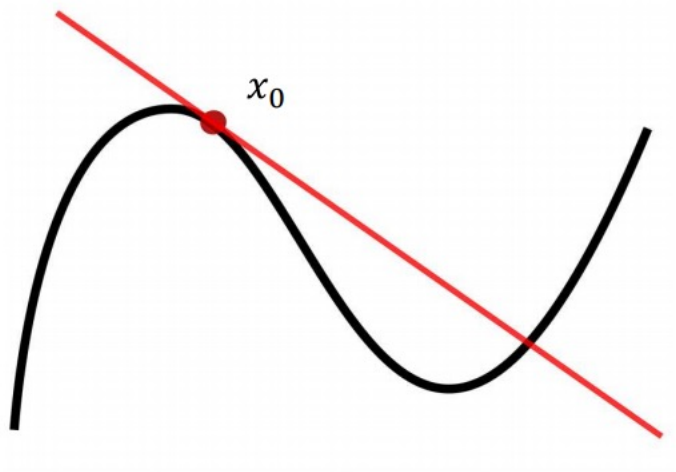
- Вектор градиента функции потерь обозначают $grad\ Q$ или ∇Q

Метод градиентного спуска



Производная

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

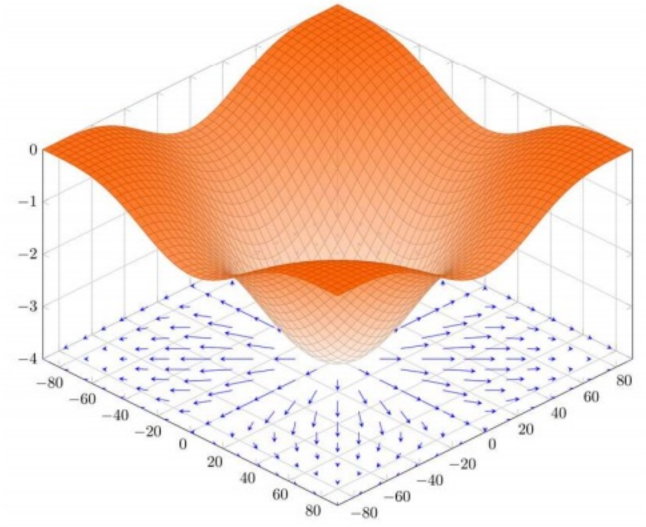


Градиент

- Градиент – вектор частных производных

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d}\right)$$

- У градиента есть важное свойство!
- Зафиксируем точку x_0
- В какую сторону функция быстрее всего растет?



Градиентный спуск

Повторять до сходимости:

$$w^t = w^{t-1} - \eta \nabla Q(w^{t-1})$$

Градиент в предыдущей точке

Новая точка

Длина шага

Метод градиентного спуска

На каждом шаге (на каждой итерации метода) движемся в сторону антиградиента Функции потерь!

- Инициализируем веса $w_0^{(0)}, w_1^{(1)}, w_2^{(0)}, \dots, w_n^{(0)}$

- На каждом следующем шаге обновляем веса, сдвигаясь в направлении антиградиента функции потерь Q:

$$w_0^{(k)} = w_0^{(k-1)} - \nabla Q\left(w_0^{(k-1)}\right),$$
$$w_1^{(k)} = w_1^{(k-1)} - \nabla Q\left(w_1^{(k-1)}\right),$$
$$\dots$$
$$w_n^{(k)} = w_n^{(k-1)} - \nabla Q\left(w_n^{(k-1)}\right),$$

Градиентный спуск

Останавливаем процесс, если

$$||w^t - w^{t-1}|| < \varepsilon$$

Другой вариант;

$$||Q(w^t) - Q(w^{t-1})|| < \varepsilon$$

Другой вариант:

$||\nabla Q(w^t)|| < \varepsilon$

Переменная длина шага

$$w^t = w^{t-1} - \eta_t \nabla Q(w^{t-1})$$

- Длину шага можно менять в зависимости от шага
- Например: $\eta_t = \frac{1}{t}$
- Шаг наискорейшего спуска:
  $\eta_t = argmin_n Q(w^t) = argmin_\eta Q(w^{t-1} - \eta \nabla Q(w^{t-1}))$

Оценка градиента

$$Q(w) = \frac{1}{l}\sum_{i=1}^{l} L(y_i, a(x_i))$$

- Градиент:

$$\nabla Q(w) = \frac{1}{l}\sum_{i=1}^{l} \nabla L(y_i, a(x_i))$$

- Может, оценить градиент одним слагаемым?

$$\nabla Q(w) \approx \frac{1}{l}\sum_{i=1}^{l} \nabla L(y_i, a(x_i))$$

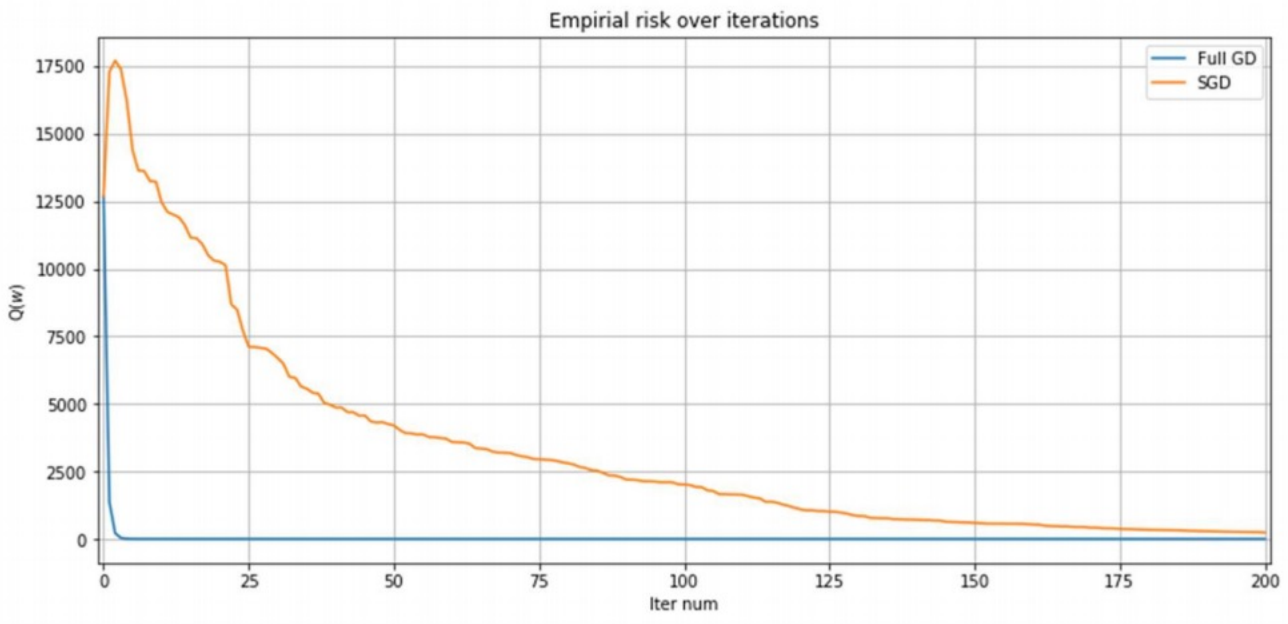
Стохастический градиентный спуск

1. Начальное приближение: w^0

2. Повторять, каждый раз выбирая случайный обьект $i_t : w^t = w^{t-1} - \eta_t \nabla L(y_{i_t,} a(x_{i_t})$

3. Останавливаемся, если
$$\left|\left|w^t - w^{t-1}\right|\right| < \varepsilon$$

Стохастический градиентный спуск



Стохастический градиентный спуск

1. Начальное приближение: w^0

2. Повторять, каждый раз выбирая m случайный обьект $i_1 , .., i_m$:
   $w^t = w^{t-1} - \eta_t \frac{1}{m} \sum_{j=1}^{m} \nabla L(y_{i_j}, a\left(x_{i_j}\right))$
3. Останавливаемся, если
   $$\left|\left|w^t - w^{t-1}\right|\right| < \varepsilon$$

Модель линейной классификации

Классификация

- $Y = \{-1, +1\}$
- -1 – отрицательный класс
- +1 – положительный класс
- Алгоритм $a(x)$ должен возвращать одно из двух чисел

Линейная регрессия

$a(x) = w_0 + \sum_{j=1}^{d} w_j x_j$ - выдает вещественное число

Линейный классификатор

$a(x) = sign(w_0 + \sum_{j=1}^{d} w_j x_j)$ - выдает знак вещественного числа

Свободный
коэффицент

веса

признаки

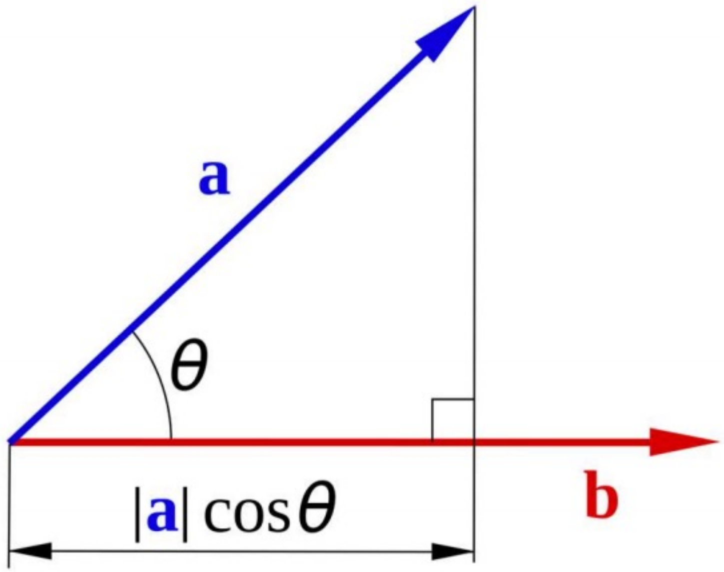
Линейный классификатор

$a(x) = sign\left(\sum_{j=1}^{d} w_j x_j\right) = sign < w, x >$ добавим единичный признак

Геометрия

Скалярное произведение :

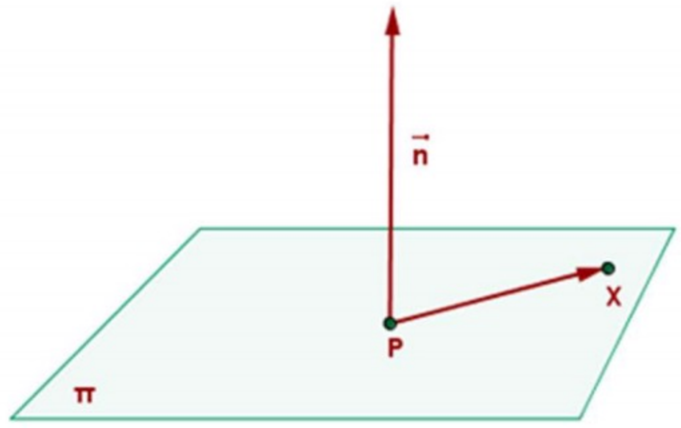
$$< a, b > = |a||b|\cos(\theta)$$



Геометрия

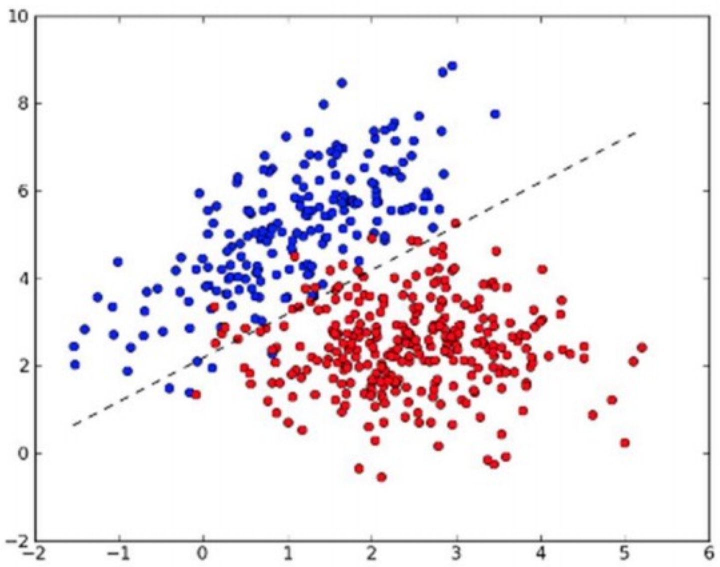
Уравнение гиперплоскости:

$$< w, x > = 0$$



Геометрия

- Линейный классификатор проводит гиперплоскость
- $< w, x > < 0$ – обьект слева от нее
- $< w, x > > 0$ – обьект справа от нее



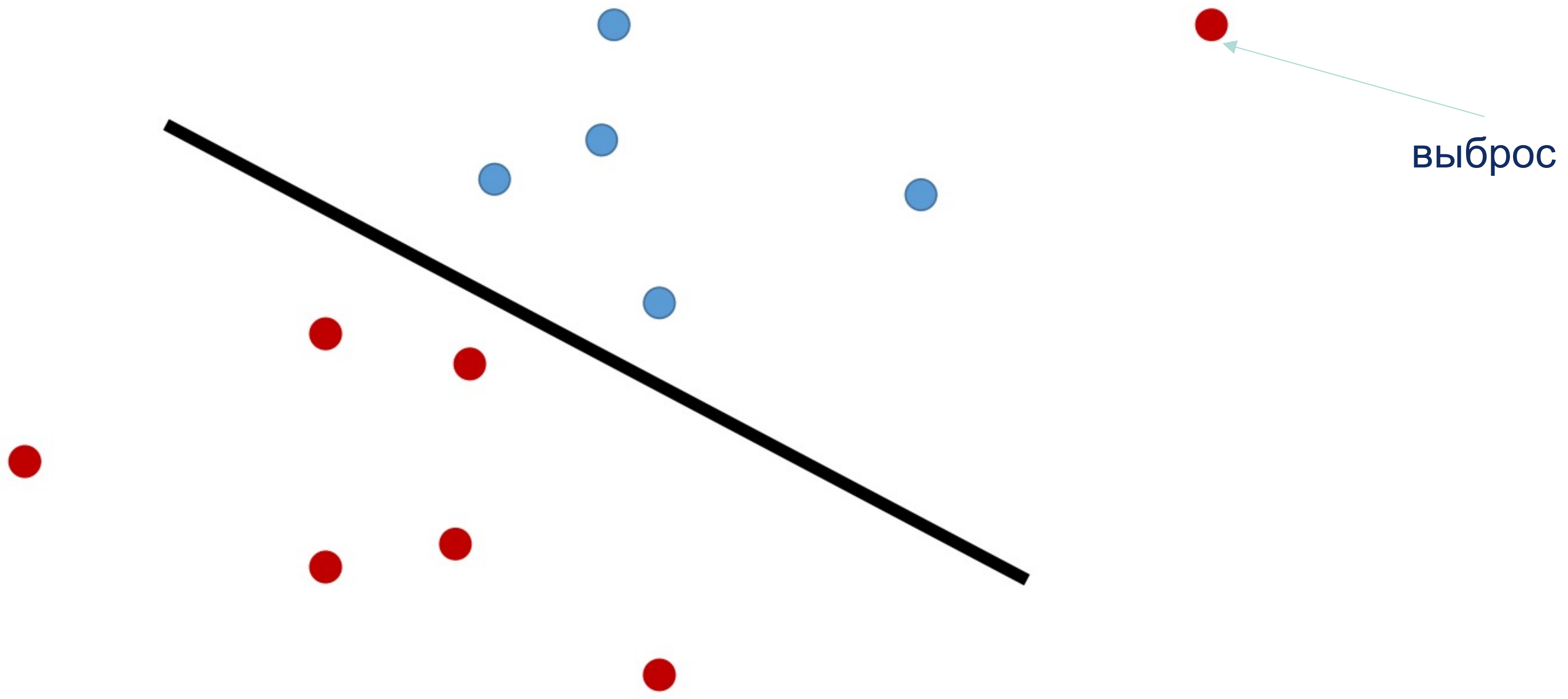
Геометрия

- Расстояние от точки до гиперплоскости $< w, x > = 0$:

$$\frac{| < w, x > |}{||w||} = \frac{| < w, x > |}{\sqrt{< w, w >}}$$

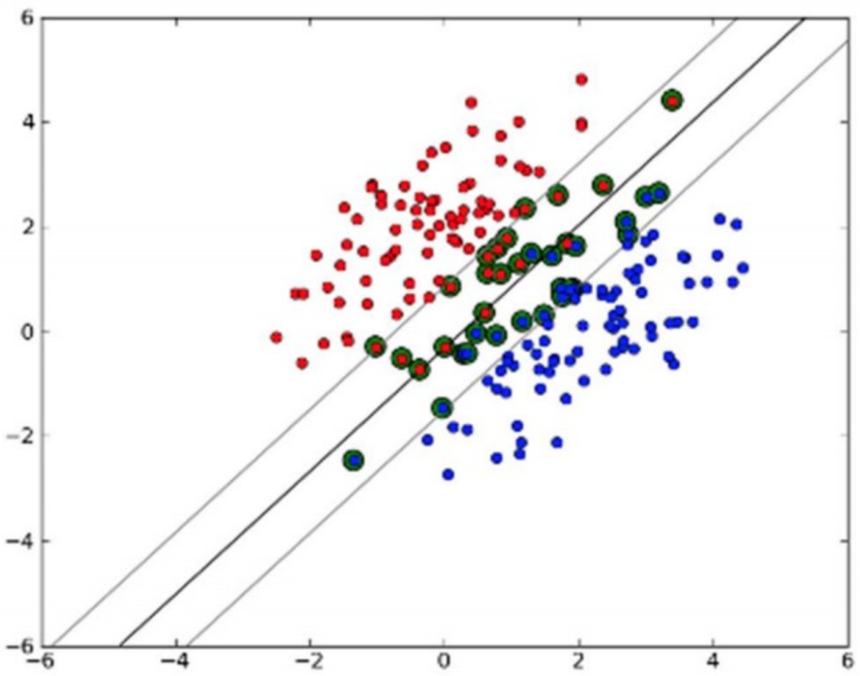
- Чем больше $< w, x >$, тем дальеш обьект от разделяющий гиперплоскости

Геометрия



Отступ

- $M_i = y_i < w, x_i >$
- $M_i > 0$ - классификатор дает верный ответ
- $M_i < 0$ - классификатор ошибается
- Чем дальше отступ от нуля, тем больше уверенности



Порог

$$a(x) = sign(< w, x > -t)$$

t - порог классификатора

Можно подбирать для оптимизации функции потерь, отличной
От использованной при обучении

Линейный классификатор

- Линейный классификатор разделяет два класса гиперплоскостью
- Чем больше отступ по модулю, тем дальше объект от гиперплоскости
- Знак отступа говорит о корректности предсказания

Обучение линейных классификаторов

Функция потерь в классификации

- Частый выбор – бинарная функция потерь
  $L(y, a) = [a \neq y]$

- Функционал ошибки – доля ошибок (error rate)
  $Q(y, a) = \frac{1}{l} \sum_{i=1}^{l} [a(x_i) \neq y_i]$

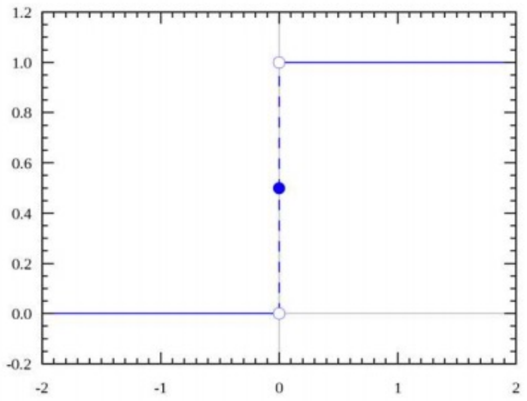
- Нередко измеряют долю верных ответов (accuracy):
  $Q(y, a) = \frac{1}{l} \sum_{i=1}^{l} [a(x_i) = y_i]$

Доля ошибок для линейного классификатора

- Функционал ошибки:

$$Q(w, X) = \frac{1}{l}\sum_{i=1}^{l}[sign(< w, x_i >) \neq y_i]$$

- Индикатор – недифференцируемая функция



Отступы для линейного классификатора

- Функционал ошибки:

$$Q(w, X) = \frac{1}{l}\sum_{i=1}^{l}[sign(< w, x_i >) \neq y_i]$$

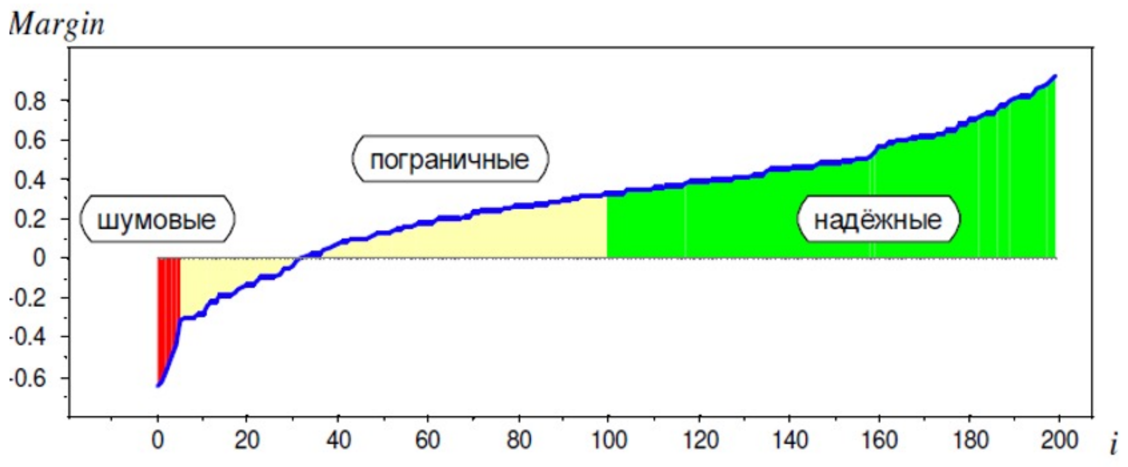
- Альтернативная запись:

$$Q(w, X) = \frac{1}{l}\sum_{i=1}^{l}[y_i < w, x_i > < 0]$$

$y_i < w, x_i > = M_i$ - отступ

Отступы (MARGIN)

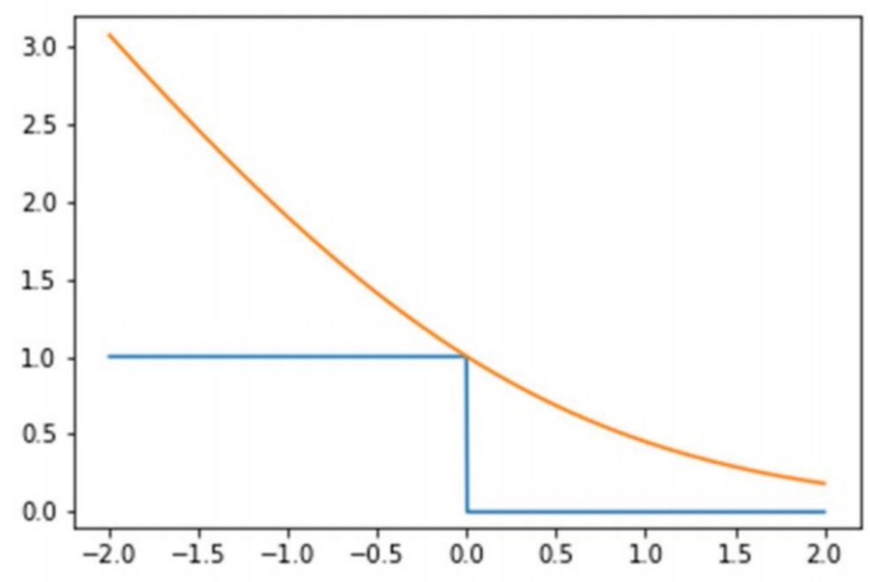
Ранжирование обьектов по возрастанию отступа:



Верхняя оценка

$L(M) = [M < 0] \leq L^{\sim}(M)$

Оценим сверху дифференцируемой функцией



Верхняя оценка

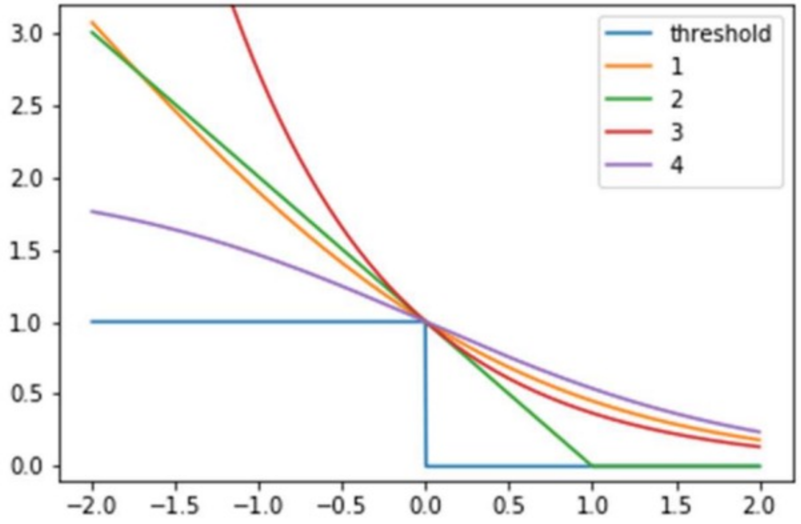
$$0 \leq \frac{1}{l}\sum_{i=1}^{l}[y_i < w, x_i > < 0] \leq \frac{1}{l}\sum_{i=1}^{l} L^{\sim}(y_i < w, x_i >) \rightarrow \min\_w$$

Минимизируем верхнюю оценку

Надеемся, что она прижмет долю ошибок к нулю

Примеры верхних оценок

- $L^{\sim}(M) = \log(1 + e^{-M})$ – логистическая
- $L^{\sim}(M) = \max(0, 1 - M)$ – кусочно – линейная
- $L^{\sim}(M) = e^{-M}$ – экспоненциальная
- $L^{\sim}(M) = \frac{2}{1+e^{M}}$ - сигмоидная



Пример обучения

Выбираем логистическую функцию потерь:

$$Q\tilde{}(w, X) = \frac{1}{l}\sum_{i=1}^{l} \log(1 + \exp(-y_i < w, x_i >)) \to \min\_w$$

Вычисляем градиент:

$$\nabla_w Q\tilde{}(w, X) = -\frac{1}{\mathrm{l}}\sum_{i=1}^{l} \frac{y_i x_i}{1+\exp(y_i<w,x_i>)}$$

Пример обучения

Делаем градиентный спуск

$$w^{(t)} = w^{(t-1)} + \eta \frac{1}{l} \sum_{i=1}^{l} \frac{y_i x_i}{1+\exp(y_i \langle w, x_i \rangle)}$$

Пример регуляризации

$$\frac{1}{l}\sum_{i=1}^{l}\log(1+\exp(-y_i < w, x_i >)) + \lambda\left|\left|w\right|\right|^2 \to \min\_w$$

Полностью аналогично линейной регрессии

Важно не накладывать регуляризацию на свободный коэффицент

https://medium.com/@shrutijadon10104776/why-we-dont-use-bias-in-regularization-5a86905dfcd6

Метрики качества классификации

Качество классификации

- Доля неправильных ответов:

$$\frac{1}{l}\sum_{i=1}^{l}[a(x_i) \neq y_i]$$

- Доля правильных ответов:

$$\frac{1}{l}\sum_{i=1}^{l}[a(x_i) = y_i]$$

Несбалансированные выборки

- Несбалансированная выборка — объектов одного класса существенного больше
- Пример: предсказание кликов по рекламе
- Пример: медицинская диагностика
- Пример: предсказание оттока клиентов
- Пример: специализированный поиск

Несбалансированные выборки

- Пример:
  - Класс -1 : 950 обьектов
  - Класс +1 : 50 обьектов
  - $a(x) = -1$
  - Доля правильных ответов : 0.95
  - Почему результат нас не устраивает?
  - Модель не несет экономической ценности
  - Цены ошибок неравнозначны

Несбалансированные выборки

- q_0 - доля обьектов самого крупного класса
- Для разных алгоритмов :

$$accuracy \in [q_0, 1]$$

- Если получили большой accuracy — посмотрите на баланс классов

Улучшение метрики

- Два алгоритма
- Доли правильных ответов : r_1 и r_2
- Абсолютное улучшение: $r_2 - r_1$
- Относительное отношение: $\frac{r_2 - r_1}{r_1}$

Улучшение метрики

- $r_1 = 0.8$,
- $r_2 = 0.9$,
- $\frac{r_2 - r_1}{r_1} = 12.5\%$

- $r_1 = 0.5$,
- $r_2 = 0.75$,
- $\frac{r_2 - r_1}{r_1} = 50\%$

- $r_1 = 0.001$,
- $r_2 = 0.01$,
- $\frac{r_2 - r_1}{r_1} = 900\%$

Цены ошибок

Пример: кредитный скоринг

- Что хуже?
    - Выдать кредит «плохому» клиенту
    - Не выдать кредит «хорошему» клиенту

- Доля верных ответов не учитывает цены ошибок

Матрица ошибок

	Y = 1	Y = -1
$a(x) = 1$	True Positive (TP)	False Positive (FP)
$a(x) = -1$	False Negative (FN)	True Negative (TN0

Матрица ошибок

Модель $a_1(x)$:

	Y = 1	Y = -1
$a(x) = 1$	80	20
$a(x) = -1$	20	80

Модель $a_2(x)$:

	Y = 1	Y = -1
$a(x) = 1$	48	2
$a(x) = -1$	52	98

Точность (precision)

Можно ли доверять классификатору при $a(x) = 1$?

$$precision(a, X) = \frac{TP}{TP + FP}$$

Точность (precision)

Модель $a_1(x)$:

	Y = 1	Y = -1
$a(x) = 1$	80	20
$a(x) = -1$	20	80

Precision$(a_1, X) = 0.8$

Модель $a_2(x)$:

	Y = 1	Y = -1
$a(x) = 1$	48	2
$a(x) = -1$	52	98

Precision$(a_2, X) = 0.96$

Полнота (recall)

Как много положительных обьектов находит классификатор?

$$recall(a, X) = \frac{TP}{TP + FN}$$

Полнота (recall)

Модель $a_1(x)$:

	Y = 1	Y = -1
$a(x) = 1$	80	20
$a(x) = -1$	20	80

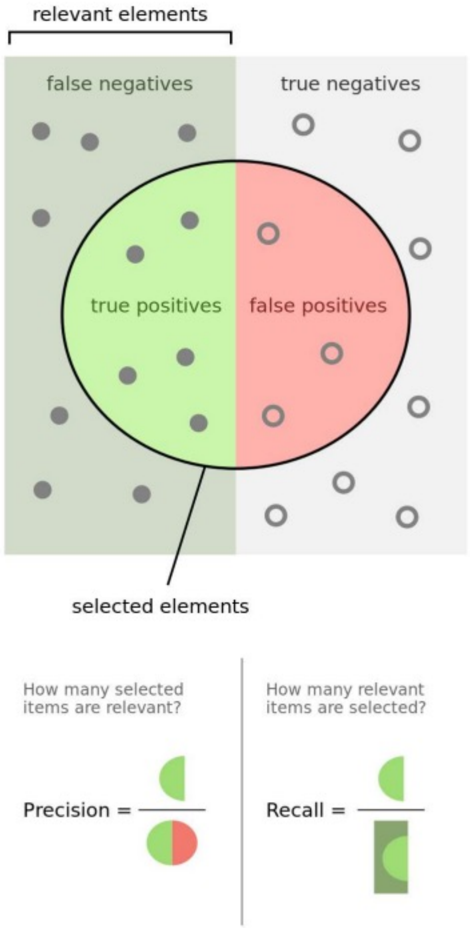
recall$(a_1, X) = 0.8$

Модель $a_2(x)$:

	Y = 1	Y = -1
$a(x) = 1$	48	2
$a(x) = -1$	52	98

recall$(a_2, X) = 0.48$

Точность и полнота



Регулируем точность и полнота

Пусть $p(x)$ – уверенность классификатора в том, что обьект x относится к классу +1, $p(x) \epsilon\ [0; 1]$

Обычно если $p(x) > 0.5$, то мы относим объект к положительному классу, а иначе – к отрицательному.

Можно изменять этот порог, то есть вместо 0.5 брать другое число из отрезка $[0; 1]$.

Регулируем точность и полнота

Пусть $p(x)$ – уверенность классификатора в том, что обьект x относится к классу +1, $p(x) \epsilon\ [0; 1]$

Обычно если $p(x) > 0.5$, то мы относим объект к положительному классу, а иначе - к отрицательному.

Можно изменять этот порог, то есть вместо 0.5 брать другое число из отрезка $[0; 1]$.

Путем изменения порога t можно регулировать точность и полноту:

- Например, если $t = 0$, то мы все объекты относим к положительному классу, то есть полнота = 1, а точность м
- При увеличении t полнота уменьшается (могут появиться объекты положительного класса, которые мы не нашли), а точность возрастет (появляются объекты положительного класса).

Антифрод

- Классификация транзакций на нормальные и мошеннические

- Высокая точность, низкая полнота:
  - Редко блокируем нормальные транзакции
  - Пропускаем много мошеннических

- Низкая точность, высокая полнота:
  - Часто блокируем нормальные транзакции
  - Редко пропускаем мошеннические

Кредитный скоринг

- Неудачных кредитов должно быть не больше 5%
- Ограничение: $precision(a, X) \geq 0.95$
- Максимизируем полноту

Медицинская диагностика

- Надо найти не менее 80% больных
- Ограничение: $recall(a, X) \geq 0.8$
- Максимизируем точность

Несбалансированные выборки

- $recall(a, X) = 0.1$
- $precision(a, X) = 0.33$
- $accuracy(a, X) = 0.99$

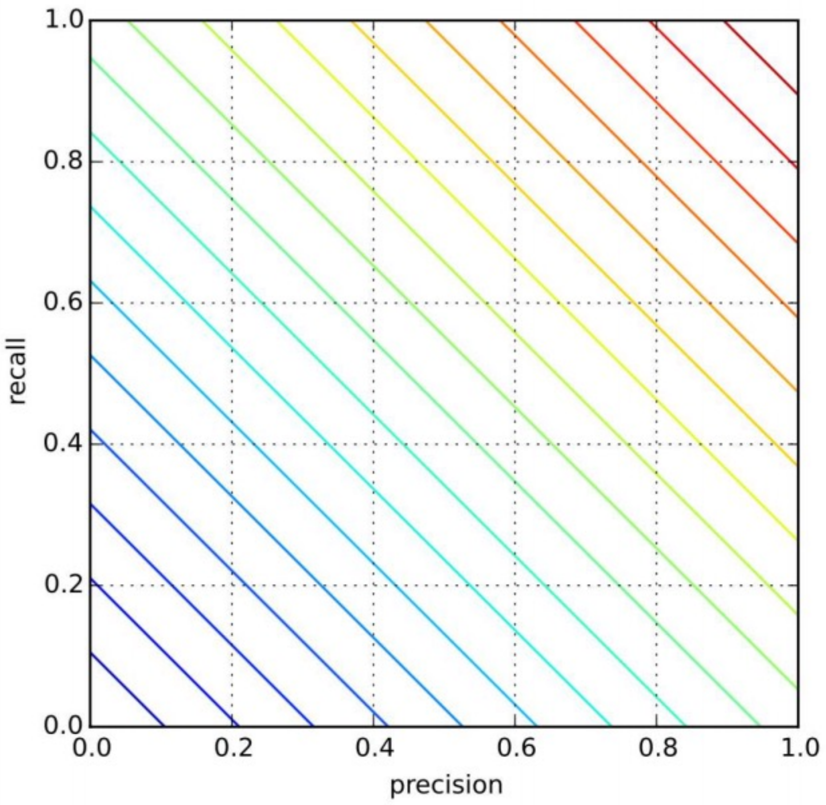
Совмещение точности и полноты

Точность и полнота

- Точность — можно ли доверять классификатору при $a(x) = 1$?
- Полнота — как много положительных объектов находит $a(x)$?
- Оптимизировать две метрики одновременно очень неудобно
- Как объединить?

Арифметическое среднее

$$A = \frac{1}{2}(precision + recall)$$

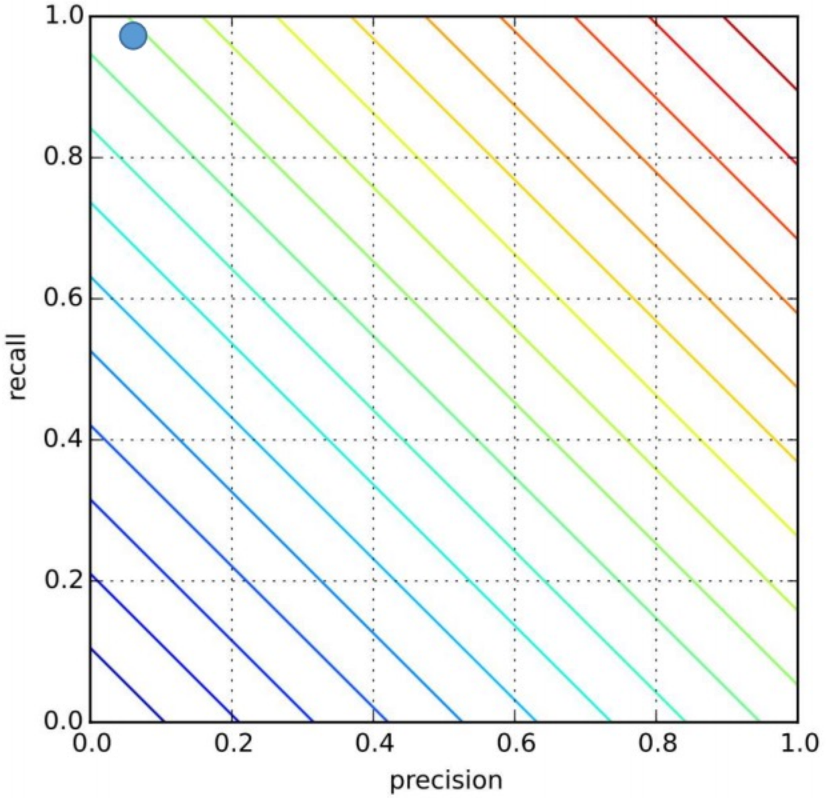


Арифметическое среднее

$$A = \frac{1}{2}(precision + recall)$$

- $precision = 0.1$
- $recall = 1$
- $A = 0.55$

- Плохой алгоритм



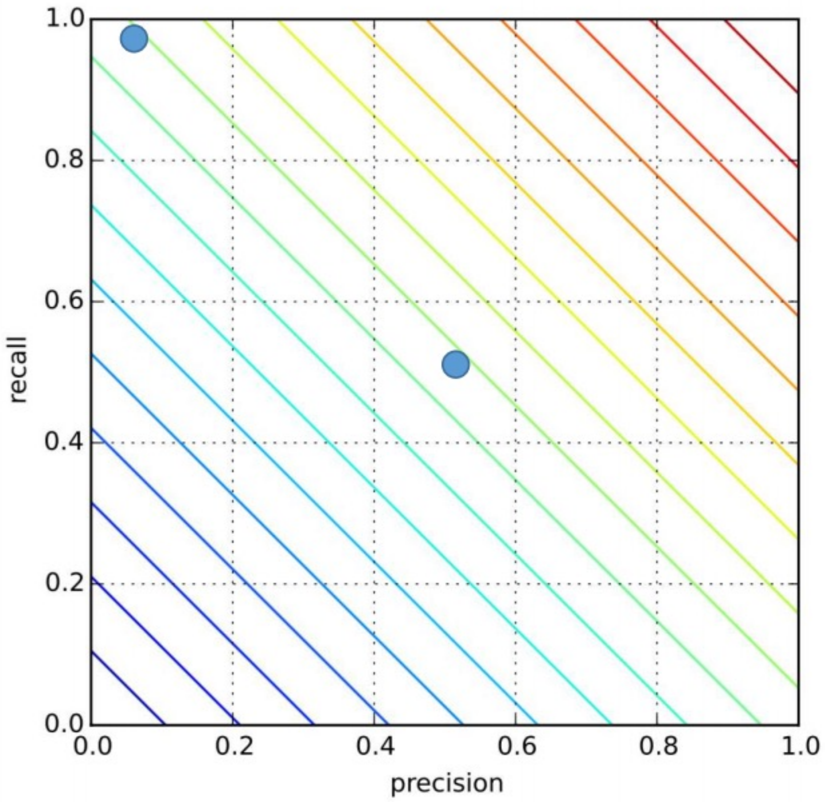
Арифметическое среднее

$$A = \frac{1}{2}(precision + recall)$$

- $precision = 0.55$
- $recall = 0.55$
- $A = 0.55$

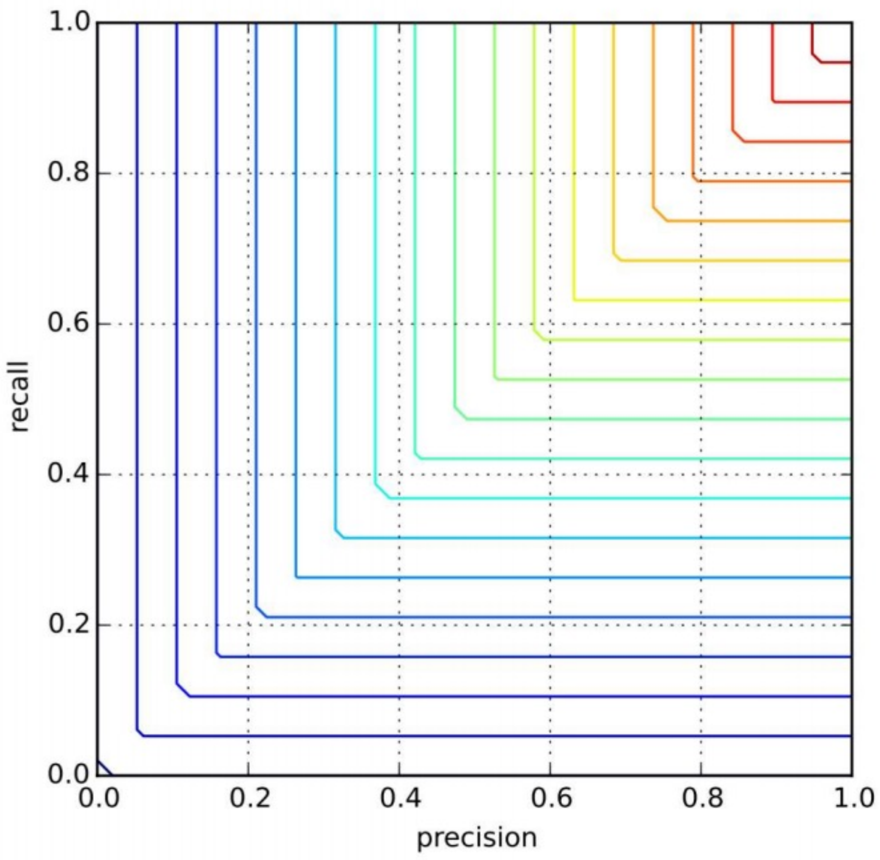
- Нормальный алгоритм

- Но качество такое же,
- как у плохого



Минимум

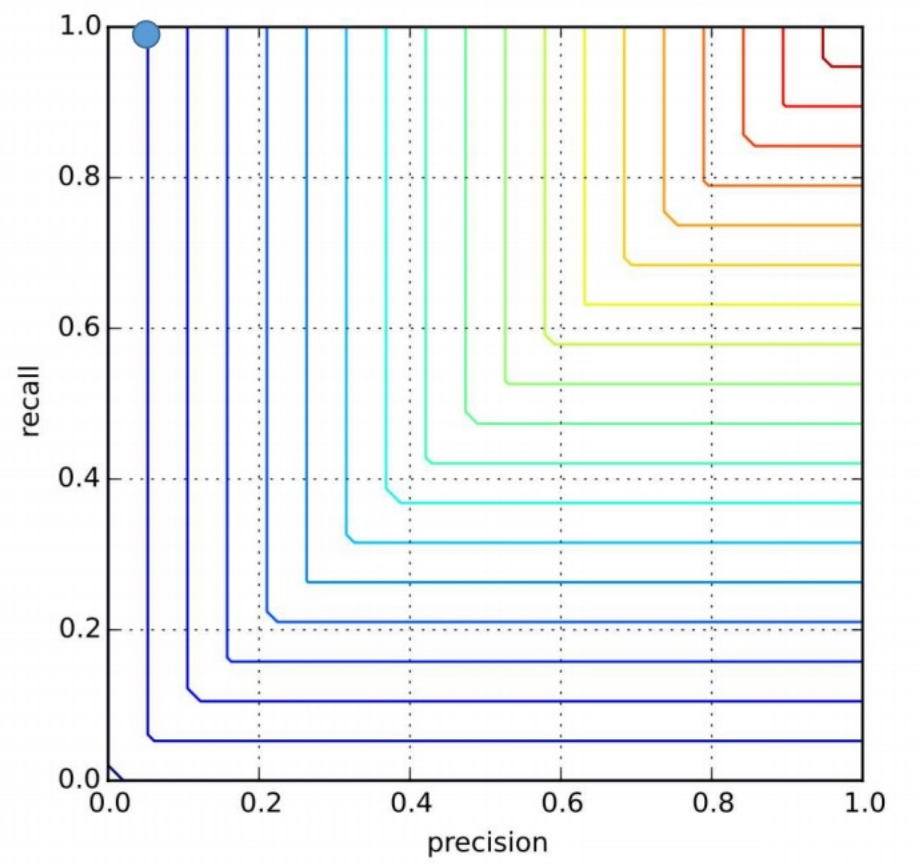
$$M = min(precision + recall)$$



Минимум

$$M = min(precision + recall)$$

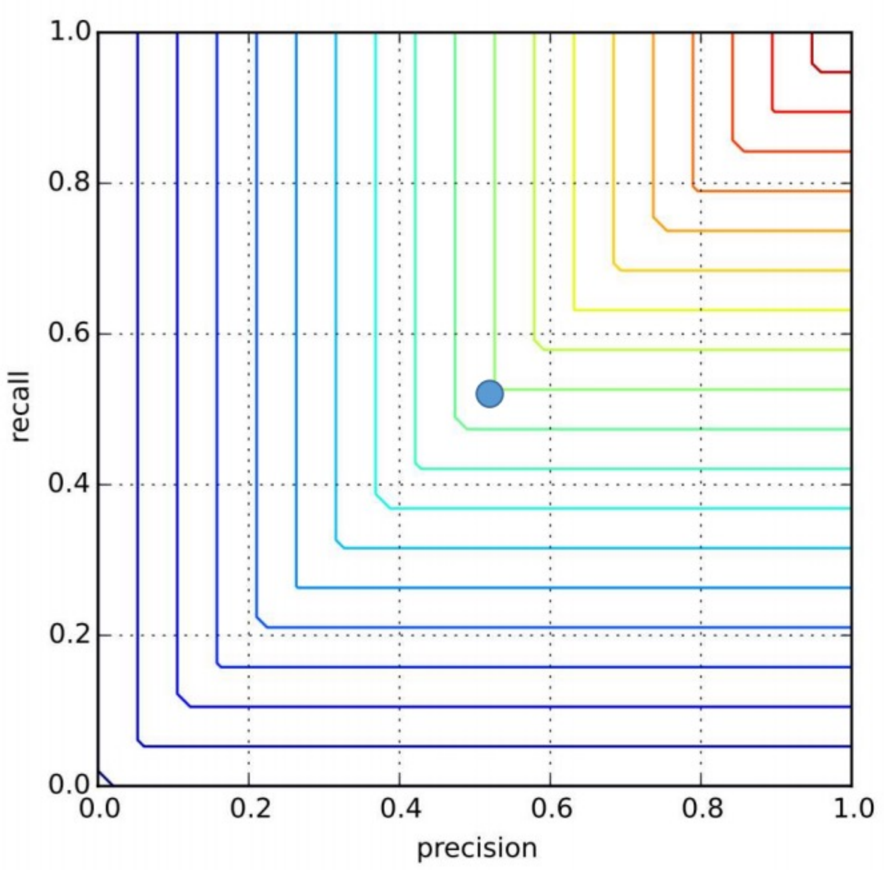
- $precision = 0.05$
- $recall = 1$
- $A = 0.05$



Минимум

$$M = min(precision + recall)$$

- $precision = 0.55$
- $recall = 0.55$
- $M = 0.55$



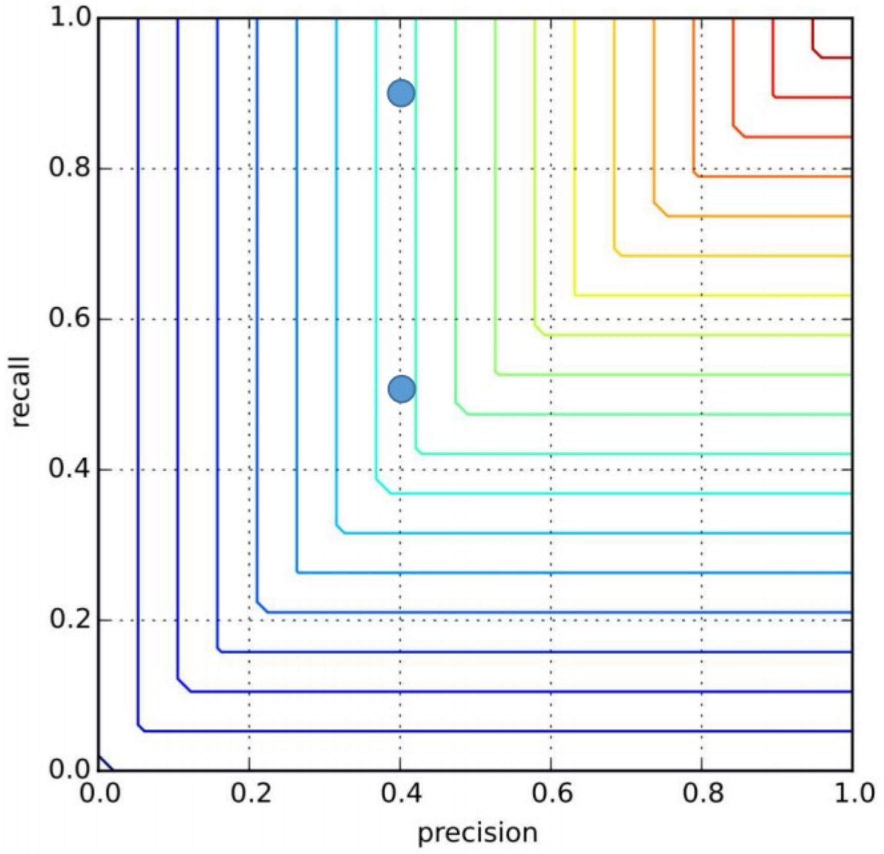
Минимум

$$\mathrm{M} = min(precision + recall)$$

- $precision = 0.4, recall = 0.5$
- $M = 0.4$

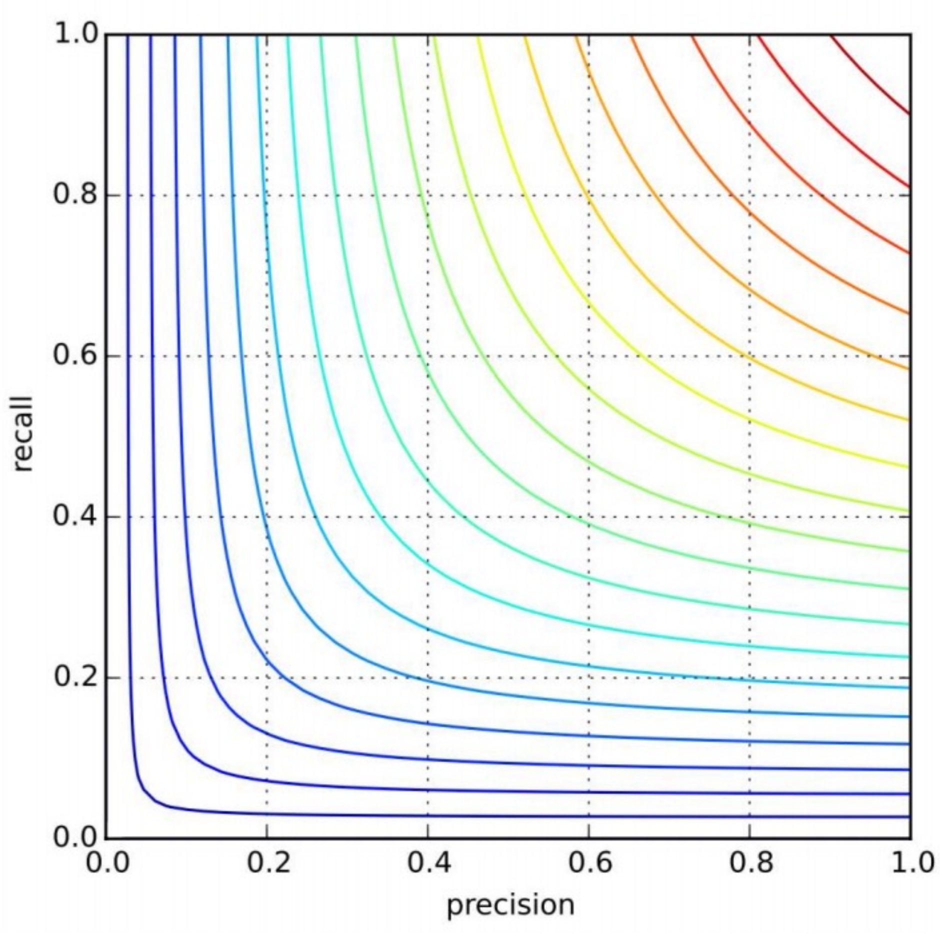
- $precision = 0.4, recall = 0.9$
- $M = 0.4$

Но второй алгоритм лучше!



F-мера

$$F = \frac{2 \ * precision \ * recall}{precision + recall}$$



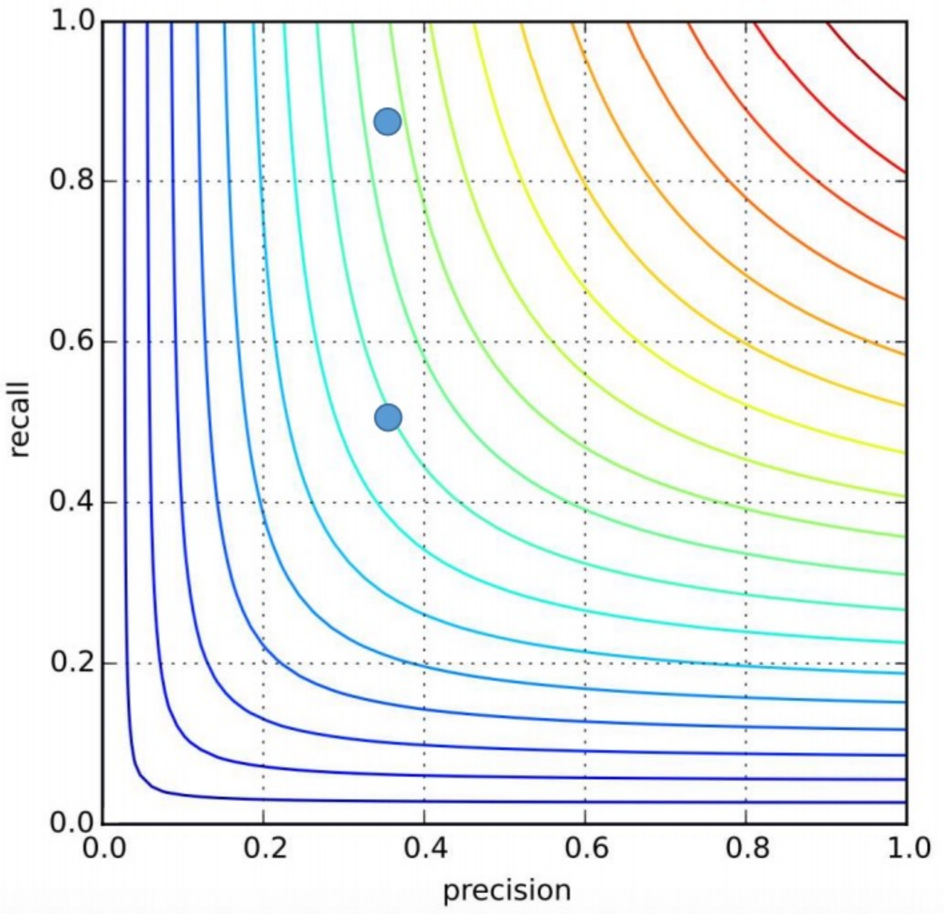
F-мера

$$F = \frac{2 \ * precision \ * recall}{precision + recall}$$

- $precision = 0.4, recall = 0.5$
- $F = 0.44$

- $precision = 0.4, recall = 0.9$
- $F = 0.55$

Работает!

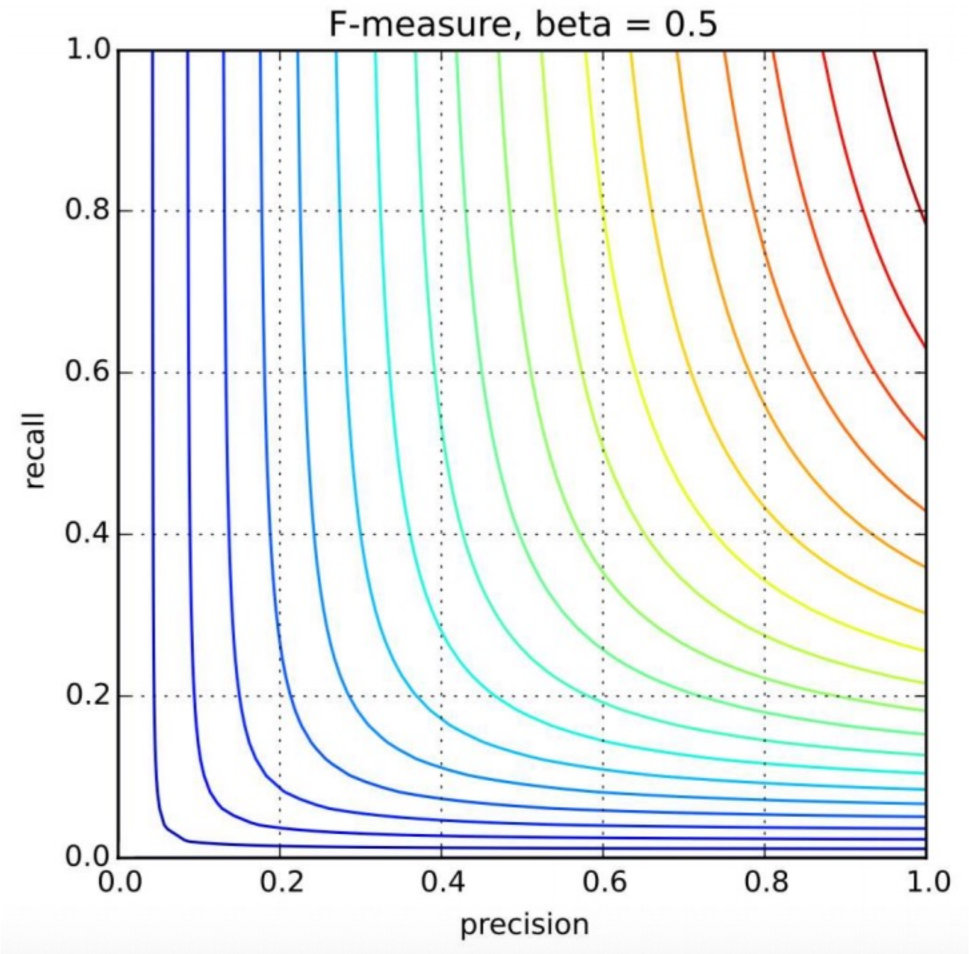


F-мера

$$F = (1 + \beta^2 \frac{*precision *recall}{\beta^2 * precision+recall})$$

$$\beta = 0.5$$

Важнее точностьс

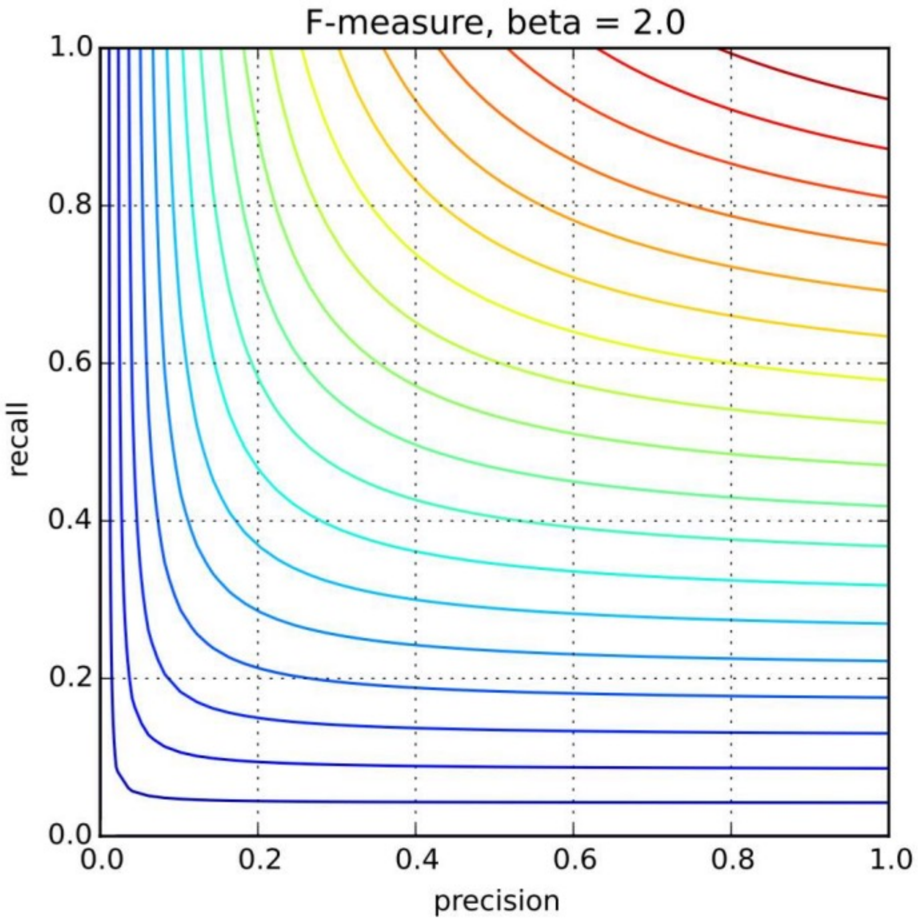


F-мера

$$F = (1 + \beta^2 \frac{*precision *recall}{\beta^2 * precision + recall})$$

$$\beta = 2$$

Важнее полнота



Метрики качества ранжирования

Классификатор

Линейный классификатор:

$a(x) = sign(< w, x > \quad -t) = 2[< w, x > \ > t] - 1$

$< w, x >$ - оценка принадлежности классу +1

Нередко $t = 0$

Оценка принадлежности

Как оценить качество $b(x)$?

Порог зависит от ограничений на точность и полноту

Оценка принадлежности

Высокий порог:

- Мало объектов относим к +1
- Точность выше
- Полнота ниже

- Низкий порог:
  - Много объектов относим к +1
  - Точность ниже
  - Полнота выше

Оценка принадлежности

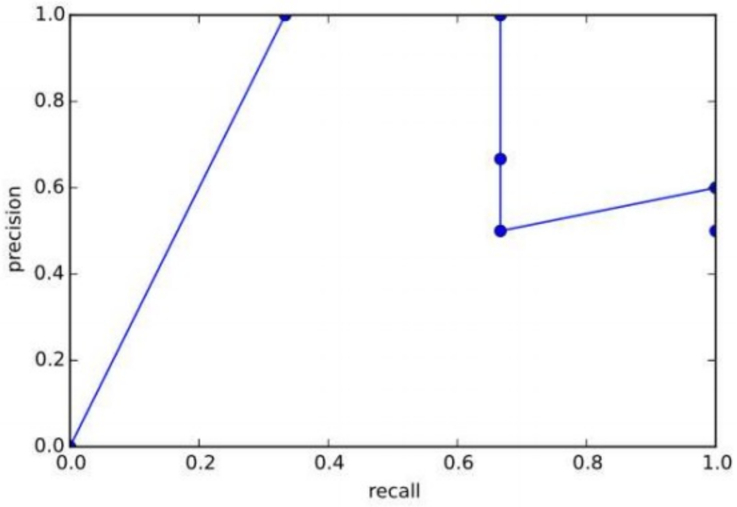
-1	-1	+1	-1	-1	-1	+1	+1	-1	+1
0.01	0.09	0.12	0.15	0.29	0.4	0.48	0.6	0.83	0.9

-1	-1	+1	-1	-1	-1	+1	+1	-1	+1
0.01	0.09	0.12	0.15	0.29	0.4	0.48	0.6	0.83	0.9

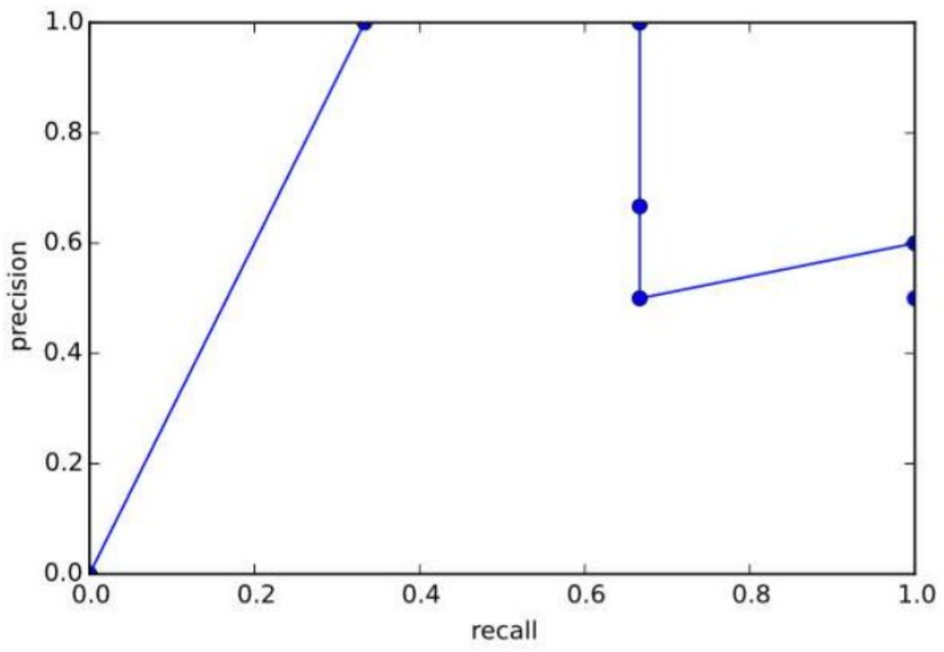
-1	-1	+1	-1	-1	-1	+1	+1	-1	+1
0.01	0.09	0.12	0.15	0.29	0.4	0.48	0.6	0.83	0.9

PR-кривая

- Кривая точности-полноты
- Ось X — полнота
- Ось Y — точность
- Точки — значения точности и
- полноты при последовательных порогах

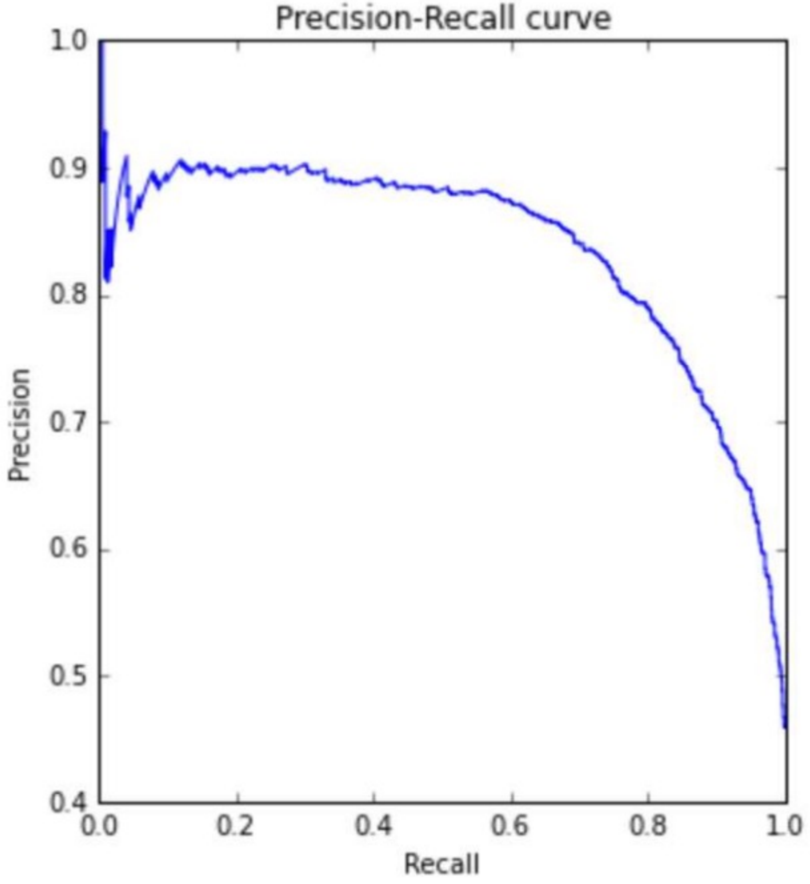


PR-кривая



$b(x)$	0.14	0.23	0.39	0.52	0.73	0.90
y	0	1	0	0	1	1

PR-кривая в реальности



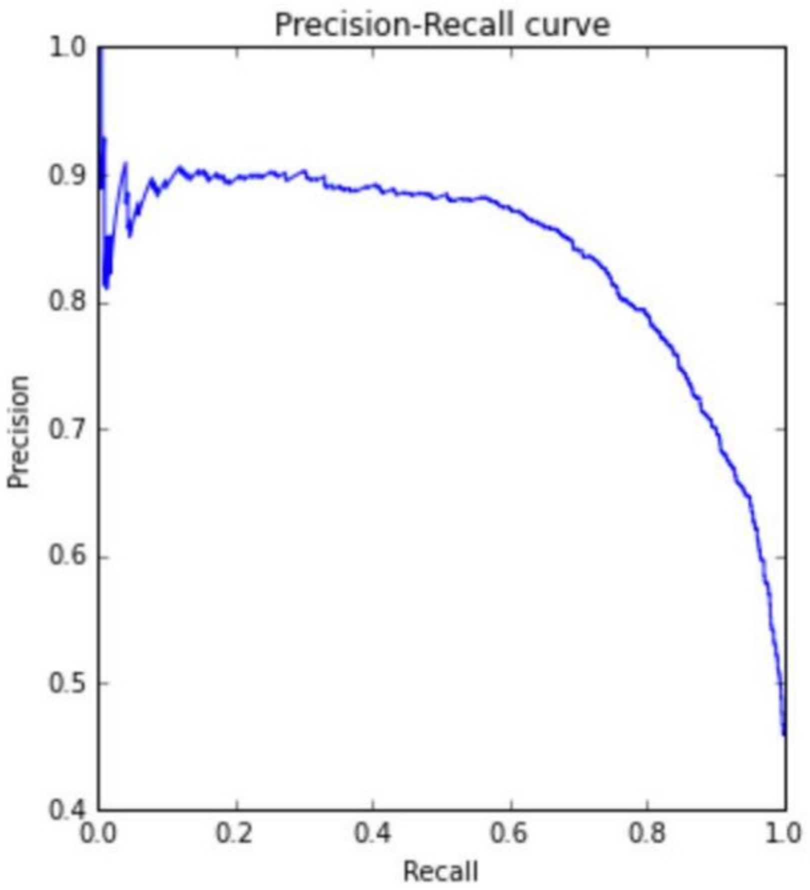
PR-кривая

Левая точка: (0, 0) или (0, 1)

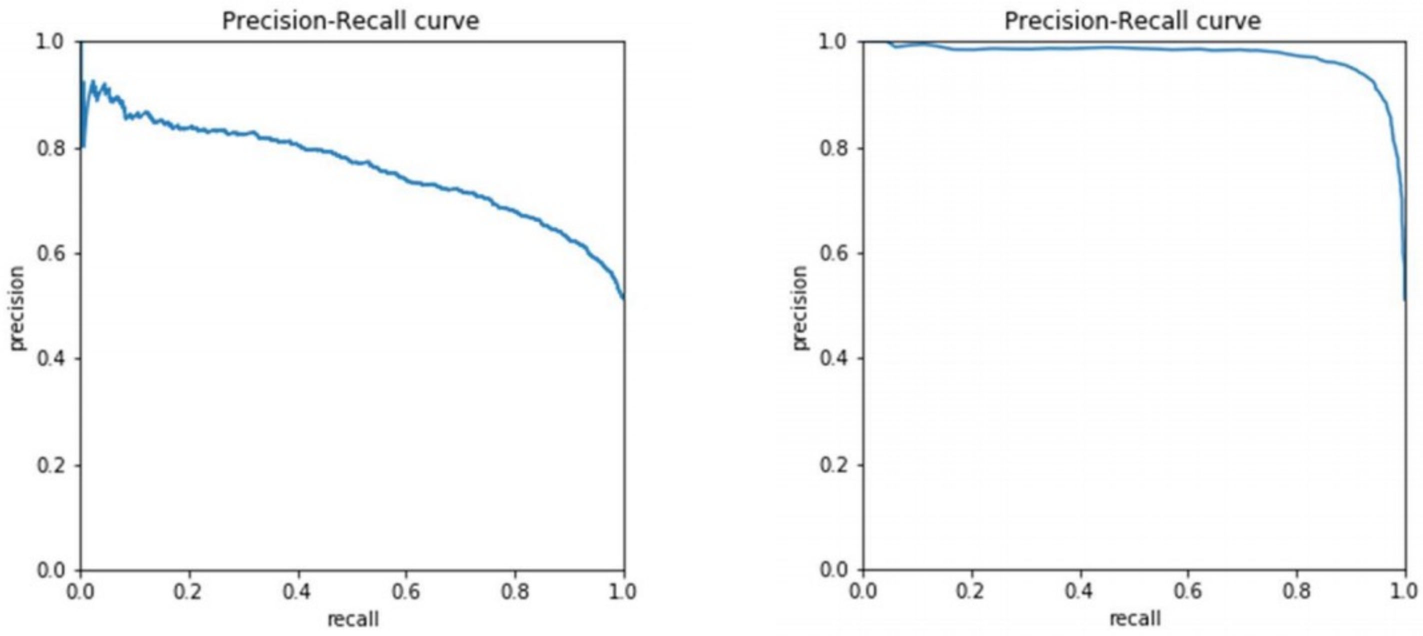
Правая точка: (1, r), r – доля положительных обьектов

Для идеального классификатора проходит Через (1, 1)

AUC-PRC – площадь под PR-кривой



PR-кривая


Precision-Recall curve
precision
recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Precision-Recall curve
precision
recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

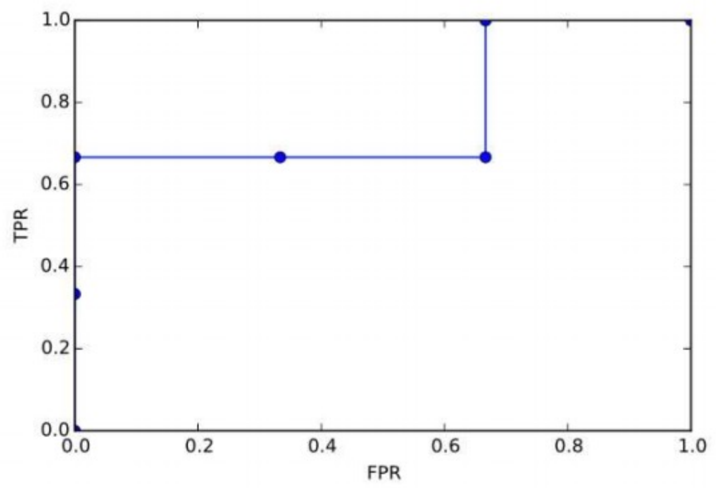
ROC-кривая

- Receiver Operating Characteristic
- Ось X — False Positive Rate

$$FPR = \frac{FP}{FP + TN}$$

- Ось Y - True Positive Rate

$$TPR = \frac{TP}{TP + FN}$$



ROC-кривая

Для каждого значения порога t вычислим:

- False Positive Rate - долю неверно принятых объектов
- True Positive Rate - долю верно принятых объектов

Кривая, состоящая из точек с координатами (FPR, TPR) для всех возможных порогов – это и есть ROC-кривая.

ROC-кривая

AUC (Area Under Curve) – площадь под ROC-кривой.

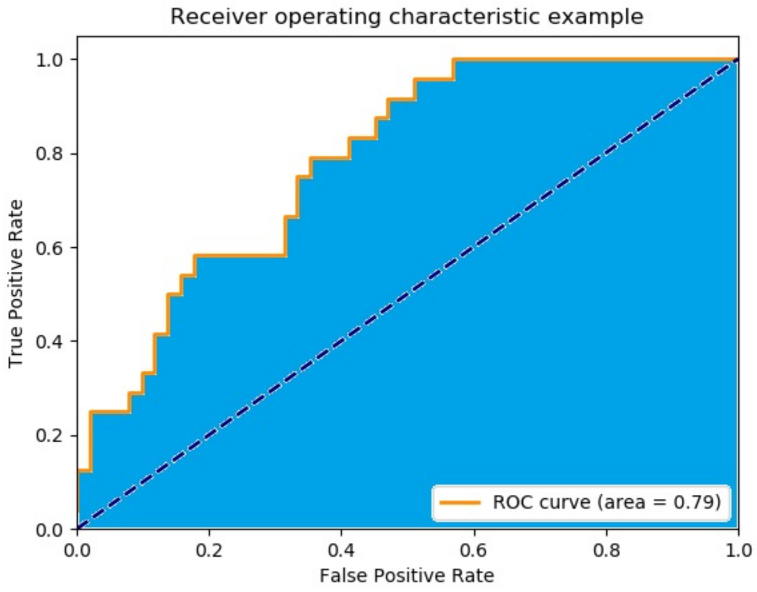
$$AUC \in [0; 1].$$

- $AUC = 1$ -

идеальная классификация

- $AUC = 0.5$ -

случайная классификация



ROC-кривая Пример

- Пусть есть выборка из 5 объектов и следующие предсказания классификатора оценки принадлежности к классу +1:

$b(x)$	0.2	0.4	0.1	0.7	0.05
y	-1	$+1$	-1	$+1$	$+1$

- Упорядочим объекты по убыванию предсказаний:

$(0.7, 0.4, 0.2, 0.1, 0.05)$

1 шаг: $\boldsymbol{t = 0.7}$, то есть

$$\boldsymbol{TPR = \frac{TP}{TP+FN}}$$

$$a(x) = [b(x) > 0.7]$$

$$\boldsymbol{FPR = \frac{FP}{FP+TN}}$$

$$TPR = \frac{0}{0+3} = 0, \ \ FPR = \frac{0}{0+2} = 0.$$

ROC-кривая Пример

- Пусть есть выборка из 5 объектов и следующие предсказания классификатора оценки принадлежности к классу +1:

$b(x)$	0.2	0.4	0.1	0.7	0.05
y	-1	$+1$	-1	$+1$	$+1$

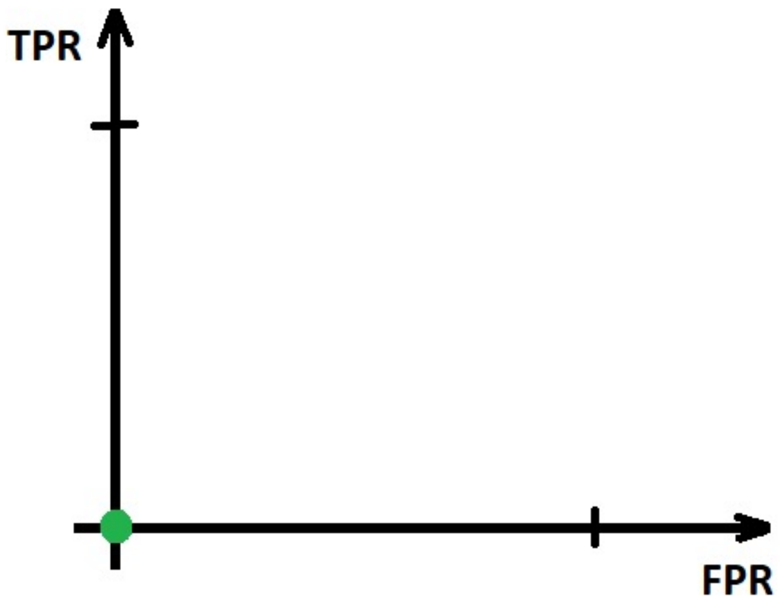
- Упорядочим объекты по убыванию предсказаний: $(0.7, 0.4, 0.2, 0.1, 0.05)$

1 шаг: $\boldsymbol{t = 0.7}$, то есть

$$a(x) = [b(x) > 0.7]$$

$$TPR = \frac{0}{0+3} = 0,$$

$$FPR = \frac{0}{0+2} = 0.$$



ROC-кривая Пример

- Оценки принадлежности к классу +1:

$b(x)$	0.2	0.4	0.1	$\mathbf{0.7}$	0.05
y	-1	$+1$	-1	$\mathbf{+1}$	$+1$

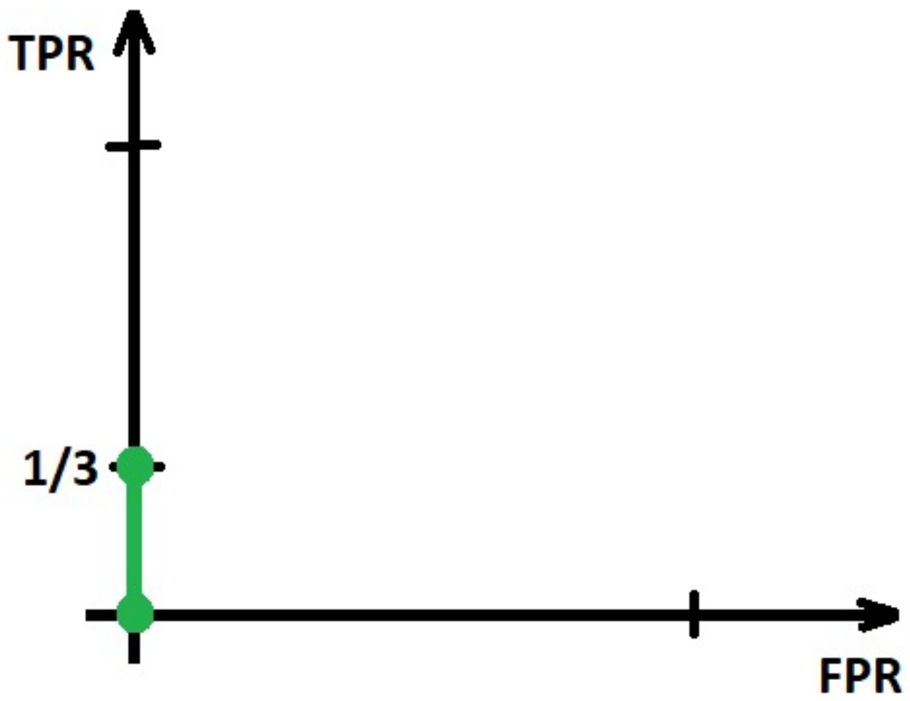
- Упорядочим объекты по

убыванию предсказаний:
$(0.7, 0.4, 0.2, 0.1, 0.05)$

2 шаг: $\boldsymbol{t = 0.4}$, то есть
$a(x) = [b(x) > 0.4]$

$TPR = \frac{1}{1+2} = \frac{1}{3}$,

$FPR = \frac{0}{0+2} = 0.$



ROC-кривая Пример

- Оценки принадлежности к классу +1:

$b(x)$	0.2	$\mathbf{0.4}$	0.1	$\mathbf{0.7}$	0.05
y	-1	$+1$	-1	$+1$	$+1$

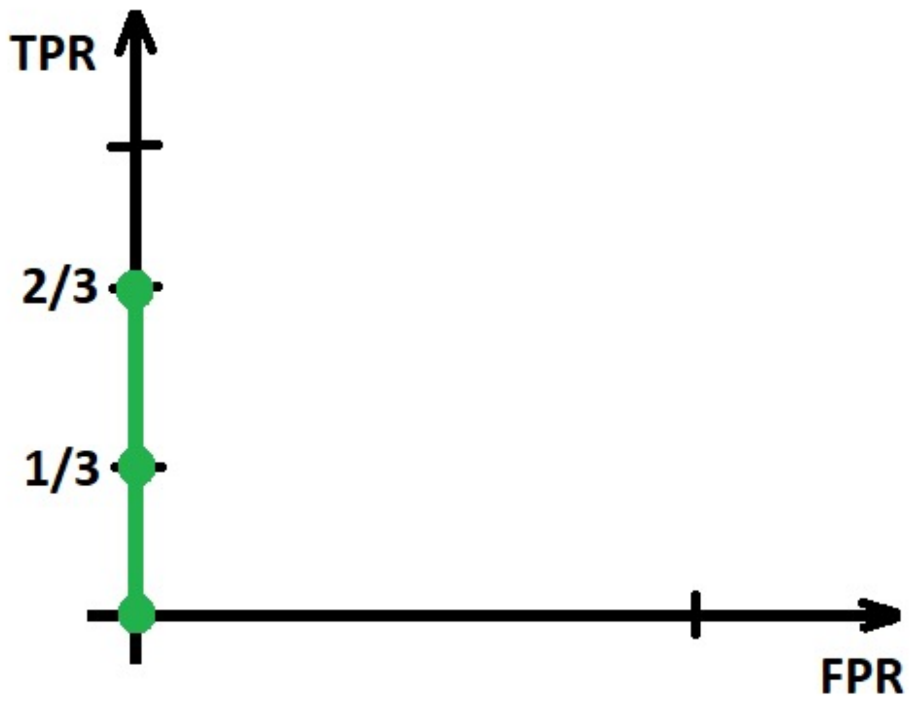
- Упорядочим объекты по

убыванию предсказаний:
$(0.7, 0.4, 0.2, 0.1, 0.05)$

3 шаг: $\boldsymbol{t = 0.2}$, то есть
$a(x) = [b(x) > 0.2]$

$TPR = \frac{2}{2+1} = \frac{2}{3}$,

$FPR = \frac{0}{0+2} = 0.$



ROC-кривая Пример

- Оценки принадлежности к классу +1:

$b(x)$	$\mathbf{0.2}$	$\mathbf{0.4}$	0.1	$\mathbf{0.7}$	0.05
y	$\mathbf{-1}$	$\mathbf{+1}$	-1	$\mathbf{+1}$	$+1$

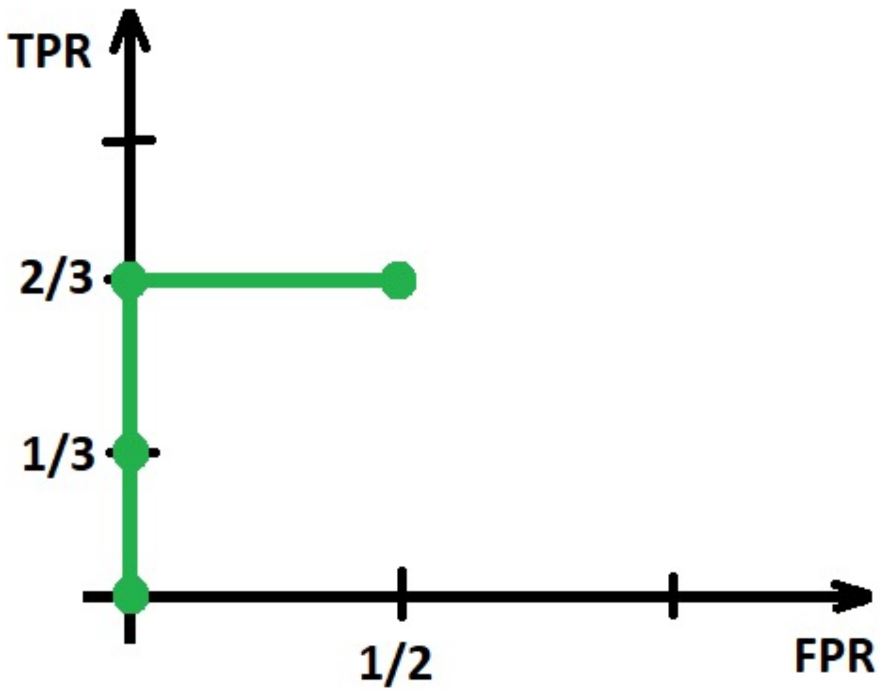
- Упорядочим объекты по

убыванию предсказаний:
$(0.7, 0.4, 0.2, 0.1, 0.05)$

4 шаг: $\boldsymbol{t = 0.1}$, то есть
$a(x) = [b(x) > 0.1]$

$TPR = \frac{2}{2+1} = \frac{2}{3}$,

$FPR = \frac{1}{1+1} = \frac{1}{2}$.



ROC-кривая Пример

- Оценки принадлежности к классу +1:

$b(x)$	$\mathbf{0.2}$	$\mathbf{0.4}$	$\mathbf{0.1}$	$\mathbf{0.7}$	0.05
y	$\mathbf{-1}$	$\mathbf{+1}$	$\mathbf{-1}$	$\mathbf{+1}$	$+1$

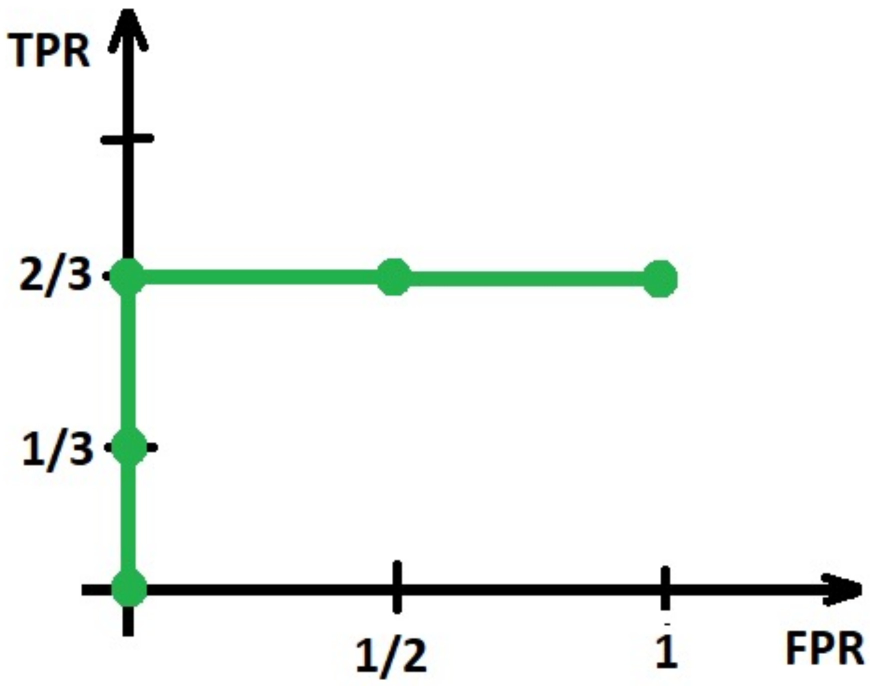
- Упорядочим объекты по

убыванию предсказаний:
$(0.7, 0.4, 0.2, 0.1, 0.05)$

5 шаг: $\boldsymbol{t = 0.05}$, то есть
$a(x) = [b(x) > 0.05]$

$TPR = \frac{2}{2+1} = \frac{2}{3}$,

$FPR = \frac{2}{2+0} = 1.$



ROC-кривая Пример

- Оценки принадлежности к классу +1:

$b(x)$	$\mathbf{0.2}$	$\mathbf{0.4}$	$\mathbf{0.1}$	$\mathbf{0.7}$	$\mathbf{0.05}$
y	$\mathbf{-1}$	$\mathbf{+1}$	$\mathbf{-1}$	$\mathbf{+1}$	$\mathbf{+1}$

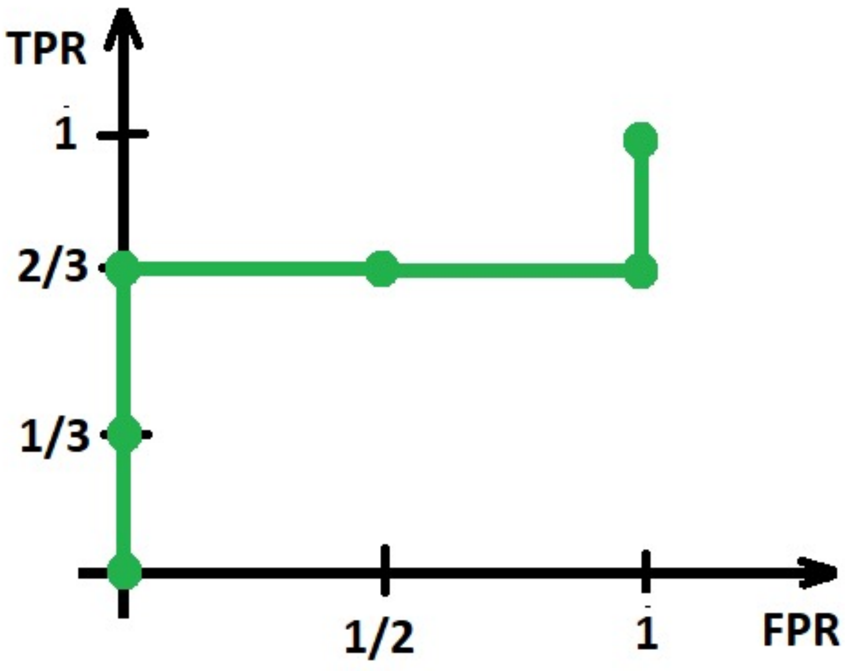
- Упорядочим объекты по

убыванию предсказаний:
$(0.7, 0.4, 0.2, 0.1, 0.05)$

5 шаг: $\boldsymbol{t = 0}$, то есть
$a(x) = [b(x) > 0]$

$$TPR = \frac{3}{3+0} = 1,$$

$$FPR = \frac{2}{2+0} = 1.$$



ROC-кривая

- Receiver Operating Characteristic

- Ось X — False Positive Rate

$$FPR = \frac{FP}{FP + TN}$$

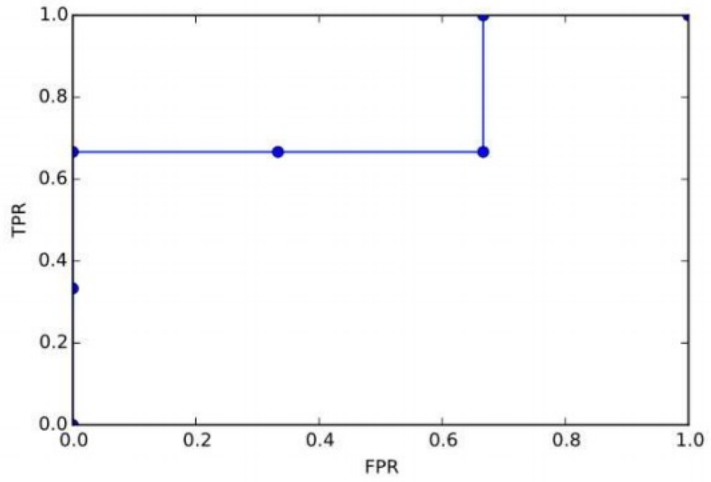
$FP + TN$ – число отрицательных обьектов

- Ось Y - True Positive Rate

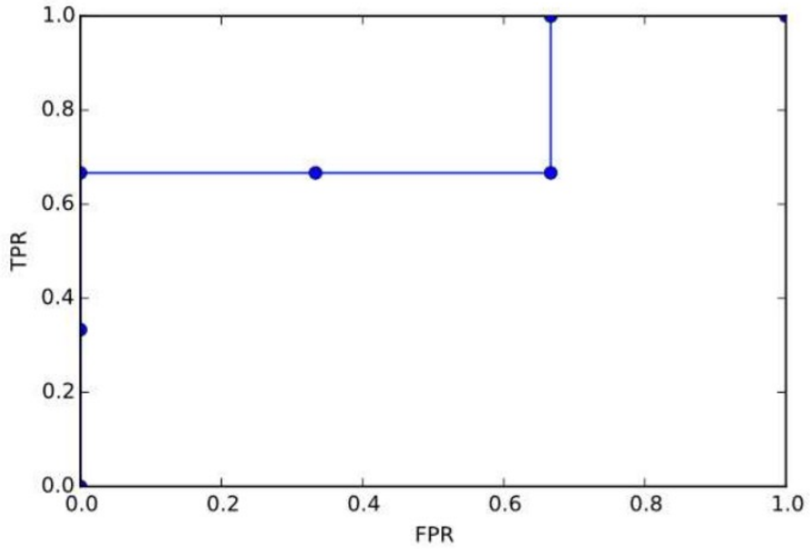
$$TPR = \frac{TP}{TP + FN}$$

Т$P + FN$ – число положительных обьектов

=

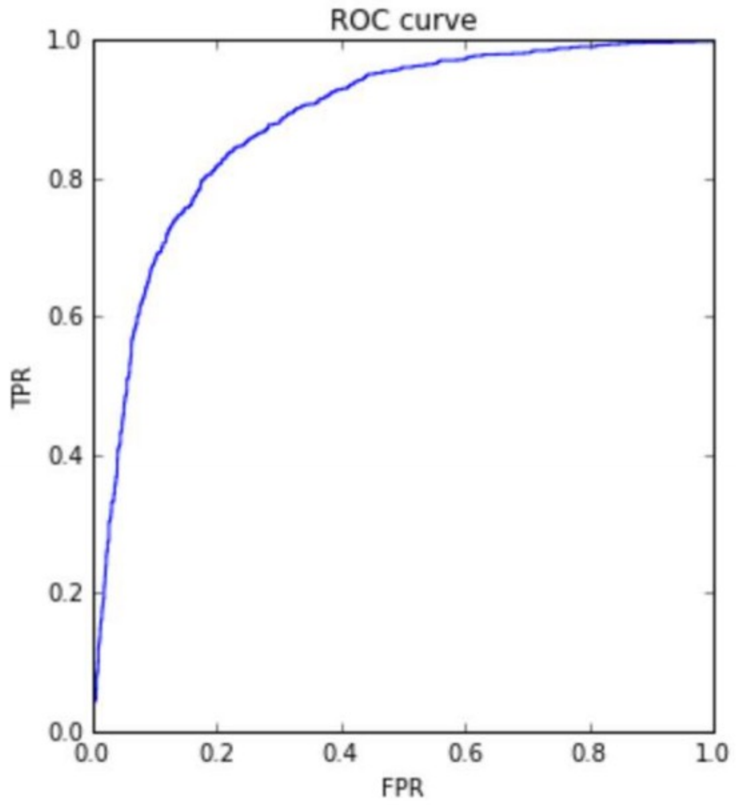


ROC-кривая



$b(x)$	0.14	0.23	0.39	0.52	0.73	0.90
y	0	1	0	0	1	1

ROC-кривая в реальности

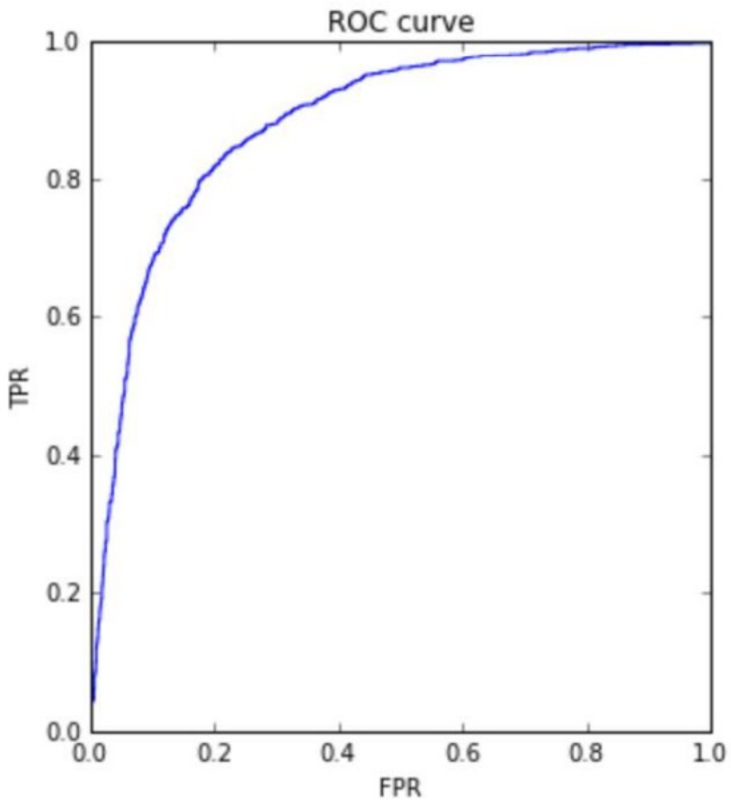


ROC-кривая

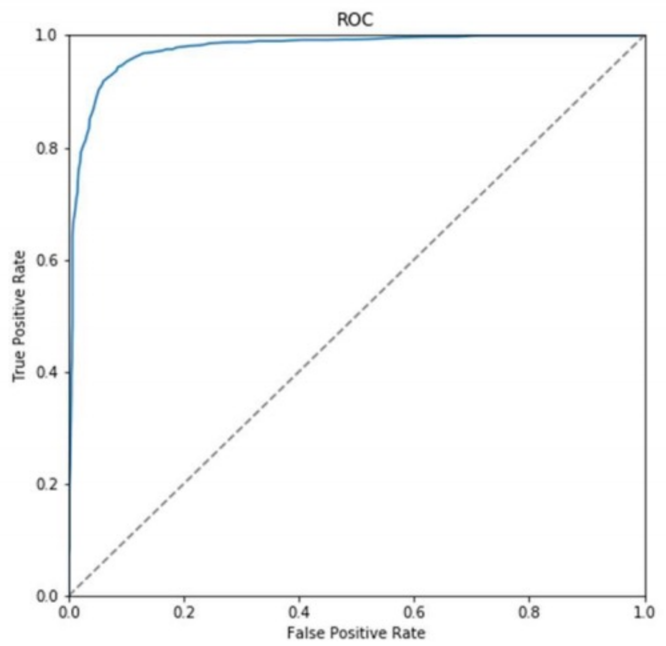
- Левая точка: (0, 0)
- Правая точка:

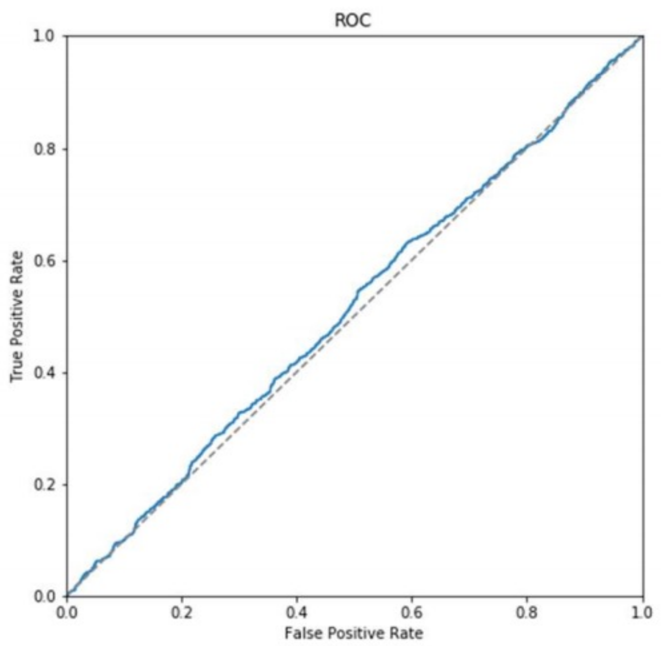
• Для идеального классификатора проходит через (0, 1)

• AUC-ROC — площадь под ROC-кривой



ROC-кривая


ROC
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0


ROC
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

AUC-ROC

$$FPR = \frac{FP}{FP + TN} \qquad TPR = \frac{TP}{TP + FN}$$

FP и TP нормируются на размеры классов

- AUC-ROC не поменяется при изменении баланса классов

- Учитывает True Negatives

- Идеальный алгоритм: $AUC - ROC = 1$

- Худший алгоритм: $AUC - ROC \approx 0.5$

AUC-PRC

$$precision = \frac{TP}{TP + FP}$$

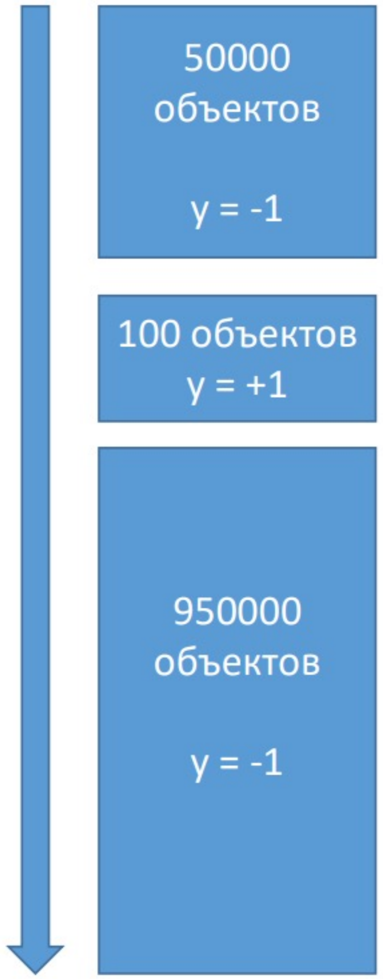
$$recall = \frac{TP}{TP + FN}$$

- Точность поменяется при изменении баланса классов
- AUC-PRC идеального алгоритма зависит от баланса классов
- Не учитывает True Negatives
- Проще интерпретировать, если выборка несбалансированная
- Лучше, если задачу надо решать в терминах точности и полноты

Пример

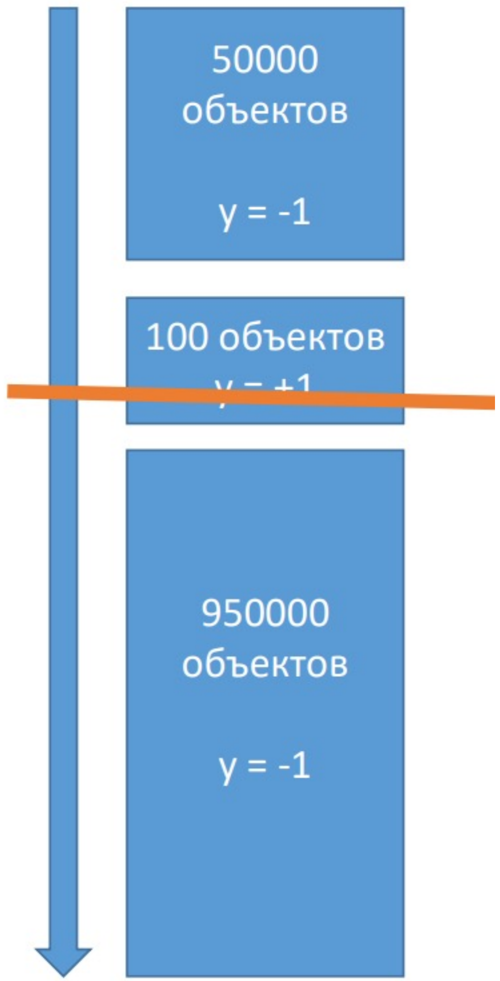
$$AUC - ROC = 0.95$$

$$AUC - PRC = 0.001$$



Пример

- Выберем конкретный классификатор
- $a(x) = 1$— 50095 объектов
- Из них FP = 50000, TP = 95
- TPR = 0.95, FPR = 0.05
- precision = 0.0019, recall = 0.95



Спасибо за внимание!

Ildar Safilo

@Ildar_Saf
irsafilo@gmail.com
https://www.linkedin.com/in/isafilo/